Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets*

Chris Bardgett† Elise Gourier‡ Markus Leippold§

March 26, 2015

Abstract

This paper studies the information content of the S&P 500 and VIX markets on the volatility of the S&P 500 returns. We estimate a flexible affine model based on a joint time series of underlying indexes and option prices on both markets. An extensive model specification analysis reveals that jumps and a stochastic level of reversion for the variance help reproduce risk-neutral distributions as well as the term structure of volatility smiles and of variance risk premia. We find that the S&P 500 and VIX derivatives prices are consistent in times of market calm but contain conflicting information on the variance during market distress.

Keywords: S&P 500 and VIX joint modeling, volatility dynamics, particle filter, variance risk premium.

JEL Codes: G12, G13, C58.

*The authors thank Yacine Aït-Sahalia, Rama Cont, Peter Christoffersen, Jérôme Detemple, Garland Durham, Damir Filipović, Andras Fulop, Michael Johannes, Loriano Mancini, Chay Ornthanalai, Chris Rogers, Ronnie Sircar, Josef Teichmann, Fabio Trojani, Anders Trolle, Alexandre Ziegler, and the seminar participants at the Gerzensee SFI PhD workshop, the Bachelier Conference 2012, the Conference of the Swiss Society for Financial Market Research 2014, the 5th International Conference of the ERCIM WG on Computing & Statistics, EFA 2013, University of Zürich, ETH Zürich, University of Aarhus and HEC Montreal for helpful comments. They further thank Jochen Krause for his implementation of the CMA-ES and DE algorithms. Financial support from the Swiss Finance Institute (SFI), Bank Vontobel, the Swiss National Science Foundation and the National Center of Competence in Research "Financial Valuation and Risk Management" is gratefully acknowledged.

†University of Zurich and Swiss Finance Institute (SFI), Plattenstrasse 14, 8032 Zürich, Switzerland; tel: (+41)-44-634-4045; Email: chris.bardgett@bf.uzh.ch.

‡ORFE Department, Princeton University, Sherrerd Hall, Princeton NJ 08544, USA; Email: egourier@princeton.edu.

§University of Zurich and Swiss Finance Institute (SFI), Plattenstrasse 14, 8032 Zürich, Switzerland; tel: (+41)-44-634-5069; Email: markus.leippold@bf.uzh.ch.

I. Introduction

Introduced by the CBOE in 1993, the VIX index non-parametrically approximates the expected future realized volatility of the S&P 500 returns over the next 30 days. Options on the VIX started trading in 2006 and, as of today, represent a much larger market than VIX futures. By definition, the VIX index, VIX options, and S&P 500 options are directly linked to the S&P 500 index and all provide valuable information on the S&P 500 returns dynamics. However, to the best of our knowledge, there has been very little effort dedicated to comparing the information these datasets contain on the distribution of the S&P 500 returns and on the trajectory of their variance process. In this paper we aim to fill this gap and study the information content of the S&P 500 and of the VIX spot and options markets.

Jointly analyzing the dynamic properties and information content of the VIX and S&P 500 option markets is a challenge. Not only do we need a set of candidate models that are flexible enough to simultaneously accommodate the stylized facts in both markets over time, but the empirical analysis of such highly nonlinear data poses a significant computational hurdle. We develop a time-consistent estimation procedure that permits us to extract information from a large and unbalanced panel of data and estimate the trajectories of the unobserved volatility of the S&P 500 returns. This methodology goes well beyond a simple calibration exercise as it makes it possible to reconcile time series data on the S&P 500 and VIX derivatives markets and consistently match the joint evolution of prices over time.

We make the following contributions to the empirical option pricing literature. First, we analyze and compare the information contained in the S&P 500 and VIX markets. We find that when markets are calm, options do not provide more information on the dynamics of volatility than the underlying S&P 500 returns and VIX levels. However, during market turmoil, our results indicate that S&P 500 options contain different information than the underlying index levels or VIX options on the trajectory of the variance factors. These findings are further supported by thorough in- and out-of-sample analyses of the ability of models estimated from one market to price options on another market. We find that the VIX index does not provide an accurate representation of the information

contained in the S&P 500 options. Similarly, the information in S&P 500 derivatives does not span the information contained in VIX derivatives, and the same holds the other way around. It is crucial to be aware of this lack of market integration when pricing and managing risks, i.e., when hedging positions on one market using a portfolio of options on the other one. We find that a joint estimation to both S&P 500 and VIX markets is required for such purposes.

Second, we perform an extensive model specification analysis. We model the S&P 500 returns using an affine jump-diffusion specification that belongs to the class of Duffie et al. (2000). This structure allows us to price S&P 500 and VIX derivatives in semi-closed form, which is essential to analyze the returns and volatility dynamics using a large dataset of options. We examine different nested models to investigate the role of each feature in explaining option prices, the risk-neutral distributions of returns, and those of the variance process. We point out the shortcomings of one-factor affine models and mitigate them thanks to a stochastic level of reversion in the volatility dynamics (also known as a stochastic central tendency).[1] Based on likelihood criteria as well as on statistical tests of the pricing errors, we find that jumps in the returns and variance processes are needed to jointly represent the index levels and derivatives prices in both markets. In particular, they make it possible to better reproduce the right tail of the variance distribution and short-maturity option prices. Furthermore, the stochastic level of reversion for the variance helps to better represent the tails of the returns' distribution and the term structure of the S&P 500 and VIX option prices.

Third, estimating the dynamics of the S&P 500 returns from an extremely large dataset of options on the two markets and for a long time series requires computationally efficient techniques that can easily deal with the features of the model, in particular the state-dependent jumps. To achieve this goal, we extend the Fourier Cosine method introduced by Fang and Oosterlee (2008) for S&P 500 options to price VIX options and adapt the Auxiliary Particle Filter of Pitt and Shephard (1999) to estimate the trajectories of unobservable processes and jumps. Accordingly, we provide an extensive toolkit for inference and diagnostics of affine option pricing models given index and option data from both the S&P 500 and VIX markets. Sequential Monte Carlo techniques have recently increased in

[1] While adding an additional factor to the Heston model increases its complexity, it has already been shown that two factors are needed to provide an accurate description of the volatility dynamics (see, e.g., Andersen et al. (2002), Alizadeh et al. (2002), Chernov et al. (2003), Christoffersen et al. (2009), Egloff et al. (2010), Todorov (2010), Kaeck and Alexander (2012), Bates (2012) and Mencía and Sentana (2013)).

popularity and have been used to estimate models, but most endeavors using this tool restrict their options dataset to near at-the-money options and as far as we know, none have used S&P 500 and VIX derivatives jointly.

Our fourth contribution is the thorough analysis of the variance risk premium. Especially for models including jumps, the identification of the risk premium's components requires a large amount of returns and options data and therefore powerful computational tools to extract the relevant information. Our estimation approach is well adapted to this task. We find that risk premia are very sensitive to jumps, in particular at the short end of the variance term structure, because a large movement in the variance process has an immediate negative impact on the payoff of a short-term variance swap. The stochastic central tendency plays a significant role in both the continuous and discontinuous parts, especially in calm markets. It is more persistent than the instantaneous variance and drives the real world medium-term expectation of variance. In contrast, the instantaneous variance reacts swiftly to changing market conditions and captures most of the sudden variance fluctuations during market turmoil.

Several papers have been published in the last years aiming to reconcile the cross-sectional information of the S&P 500 and the VIX derivatives markets by modeling them jointly. Gatheral (2008) pointed out first that even though the Heston model performs fairly well at pricing S&P 500 options, it fails to price VIX options. In fact, modeling the instantaneous volatility as a square root process leads to a VIX smile decreasing with moneyness, which is the opposite of what is observed in practice. Among the recent papers that have attempted to simultaneously reproduce the volatility smiles of S&P 500 and VIX options are Chung et al. (2011), Cont and Kokholm (2013), Song and Xiu (2012), Papanicolaou and Sircar (2013), and Bayer et al. (2013). We build on this literature by considering extensions of the Heston model that remain within the affine framework, but add more flexibility to the specifications used in the above mentioned papers. We use a special case of the general affine framework developed by Duffie et al. (2000) that includes as sub-cases the usual extensions of the Heston model encountered in the literature, for example Bates (2000), Eraker (2004) and Sepp (2008a).[2]

[2]Some studies are going in the direction of non-affine models (e.g., Jones (2003), Aït-Sahalia and Kimmel (2007), Christoffersen et al. (2010), Ferriani and Pastorello (2012), Durham (2013), Kaeck and Alexander (2012)). However,

Time-consistent estimation methods have been previously used to calibrate models to index returns and options. See, e.g., Pan (2002), Eraker (2004), Broadie et al. (2007), Christoffersen et al. (2010), Johannes et al. (2009) and Duan and Yeh (2011). However, as underlined in Ferriani and Pastorello (2012), most papers filtering information from option prices rely on one option per day or a limited set of options. Limiting the amount of data results in a computationally less intensive empirical exercise, but it ignores a large part of the information present in the markets. In contrast, in our particle filter estimation we fully exploit the richness of our dataset. Furthermore, we note that most if not all papers that consider S&P 500 and VIX options in their calibration exercise have restricted their analysis to a static one-day estimation. The resulting parameters might exhibit large variations when calibrating the model to different dates and therefore cannot be used to infer time series properties of returns and risk premia.[3]

This paper is organized as follows. In Section II, we introduce the two-factor affine jump–diffusion framework used later in the estimation. We describe the risk premium specification and derive the expression for the VIX squared as well as the pricing formula for VIX and S&P 500 options. In Section III, we describe our datasets and perform a preliminary model selection exercise based on a daily joint calibration to the S&P 500 and VIX markets. In Section IV, we detail our time series consistent estimation method. In Section V, we discuss our estimation results. Section VI concludes.

II. Theoretical framework

We first present our modeling framework, which belongs to the affine class. Then, we discuss the risk premium specification and derive the pricing formula for VIX derivatives.

A. Model specification

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space satisfying the usual assumptions, where $\mathbb{P}$ denotes the historical measure. We consider a risk-neutral measure $\mathbb{Q}$ equivalent to $\mathbb{P}$. Let $(F_t)_{t\geq 0}$ be

tractability remains an issue that is of crucial importance when it comes to calibrating a model to a long time series containing hundreds of options each day.

[3] See, e.g., Lindström et al. (2008).

the forward price of the S&P 500 index and $Y = (Y_t)_{t\geq 0} = (\log(F_t))_{t\geq 0}$ the returns. The dynamics of Y under $\mathbb{Q}$ are specified by

$$dY_t = [-\lambda^{Yv}(v_{t^-}, m_{t^-})(\theta_Z(1,0,0) - 1) - \frac{1}{2}v_{t^-}]dt + \sqrt{v_{t^-}}dW_t^Y + dJ_t^Y, \tag{1}$$

$$dv_t = \kappa_v(m_{t^-} - v_{t^-})dt + \sigma_v\sqrt{v_{t^-}}dW_t^v + dJ_t^v, \tag{2}$$

$$dm_t = \kappa_m(\theta_m - m_{t^-})dt + \sigma_m\sqrt{m_{t^-}}dW_t^m + dJ_t^m, \tag{3}$$

where W^Y, W^v, and W^m are three Brownian motions under $\mathbb{Q}$ with dependence structure

$$d\langle W^Y, W^v\rangle_t = \rho_{Y,v}dt,\ d\langle W^m, W^Y\rangle_t = 0,\ d\langle W^m, W^v\rangle_t = 0. \tag{4}$$

The process $v = (v_t)_{t\geq 0}$ is the diffusive component of the variance of S&P 500 returns. The second variance factor $m = (m_t)_{t\geq 0}$ represents the stochastic central tendency. The two processes are instantaneously uncorrelated and only interact via the drift term of v.

The processes J^Y, J^v, and J^m are finite activity jump processes driven by the point processes N^{Yv} and N^m and are defined by

$$J_t^Y = \sum_{i=1}^{N_t^{Yv}} Z_i^Y,\ J_t^v = \sum_{i=1}^{N_t^{Yv}} Z_i^v,\ J_t^m = \sum_{i=1}^{N_t^m} Z_i^m, \tag{5}$$

where Z_i^Y, Z_i^v and Z_i^m are the random jump sizes. As suggested by the price paths of the S&P 500 and VIX index, large movements in equity returns and variance are likely to occur at the same time. We therefore choose, in line with the literature, the same point process N^{Yv} to generate jumps in the asset returns and variance process v. The leverage effect is driven by the correlation between W^Y and W^v as well as the possibility of simultaneous jumps in the returns and variance. We assume

that the jump intensities depend linearly on the factor levels:[4]

$$\lambda^m(m_{t^-}) = \lambda_0^m + \lambda_1^m m_{t^-}, \tag{6}$$

$$\lambda^{Yv}(v_{t^-}, m_{t^-}) = \lambda_0^{Yv} + \lambda_1^{Yv} v_{t^-} + \lambda_2^{Yv} m_{t^-}. \tag{7}$$

Moreover, we assume that the random jump sizes are independent and identically distributed. The jump sizes in the returns are assumed to be normally distributed $\mathcal{N}(\mu_Y, \sigma_Y)$ and the jump sizes in the two volatility factors are exponentially distributed with respective means ν_v and ν_m. Let us define $Z_i = (Z_i^Y, Z_i^v, Z_i^m)^\top$, $i \in \mathbb{N}^*$. The jump sizes are characterized by their joint Laplace transform

$$\theta_Z(\phi) = \theta_{Z_1}(\phi_Y, \phi_v, \phi_m) = \mathbb{E}^{\mathbb{Q}}[\exp(\phi^\top Z_1)], \quad \phi \in \mathbb{C}^3. \tag{8}$$

This model specification implicitly defines the dynamics for the VIX. To derive its expression within our framework, we use the definition of the VIX as a finite sum of call and put prices that converges to the integral $\text{VIX}_t^2 = \frac{2}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\tau} \frac{dF_u}{F_{u^-}} - d(\ln F_u)\right]$, where τ is 30 days in annual terms.

Proposition II.1. *Under the model specification given in Equations* (1)-(8)*, the VIX squared at time t can be written as an affine deterministic function of v_t and m_t:*

$$\begin{aligned} VIX_t^2 &= \frac{1}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\tau} v_u du + 2\sum_{i=N_t^{Yv}}^{N_{t+\tau}^{Yv}} \left(e^{Z_i^Y} - 1 - Z_i^Y\right)\right], & (9)\\ &= \alpha_{VIX^2} v_t + \beta_{VIX^2} m_t + \gamma_{VIX^2}, & (10) \end{aligned}$$

where the coefficients α_{VIX^2}, β_{VIX^2} and γ_{VIX^2} are known in closed form and provided in Appendix A.

The proof of this proposition is in Appendix A.

[4]The specification of jumps is of importance. Todorov (2010), Todorov and Tauchen (2011) and Jacod and Todorov (2010) find striking evidence for co-jumps in S&P 500 returns and in the VIX. See also Eraker (2004), Broadie et al. (2007), Cont and Kokholm (2013). Bates (1996), Pan (2002) and Eraker (2004) argue in favor of using state-dependent jumps in returns, which is intuitively appealing, as jumps tend to occur more frequently when volatility increases. Using variance swaps, Aït-Sahalia et al. (2012) find that the state dependent intensity of jumps is a desirable model feature.

B. Risk premium specification

We specify the change of measure from the pricing to the historical measure so that the model dynamics keeps the same structure under $\mathbb{P}$. We separate the total equity risk premium γ_t into a diffusive contribution, which is proportional to the variance level and represents the compensation for the diffusive price risk, and a jump contribution reflecting the compensation for jump risk:

$$\gamma_t = \eta_Y v_{t^-} + \lambda^{Yv}(v_{t^-}, m_{t^-})\Big(\theta_Z^{\mathbb{P}}(1,0,0) - \theta_Z(1,0,0)\Big), \tag{11}$$

where $\theta_Z^{\mathbb{P}}$ denotes the joint Laplace transform of jump sizes under the historical measure $\mathbb{P}$. We follow Pan (2002) and Eraker (2004) and assume that the intensity of jumps is the same under $\mathbb{Q}$ and $\mathbb{P}$.[5] However, we allow the mean and volatility of the jump sizes in returns to be different under $\mathbb{Q}$ and $\mathbb{P}$.[6]

Similarly, the volatility risk premium on the two volatility factors v and m decomposes into a diffusive component and a jump component. The diffusive variance risk premium in v is proportional to the current level of variance, with coefficient of proportionality given by $\eta_v = \kappa_v - \kappa_v^{\mathbb{P}}$. The same applies to the central tendency m, for which the coefficient is defined as $\eta_m = \kappa_m - \kappa_m^{\mathbb{P}}$. For the jump part of the volatility risk premium, we allow the mean jump sizes ν_v and ν_m to be different under $\mathbb{P}$ and $\mathbb{Q}$.

For the estimation procedure, it is helpful to summarize the model in terms of $\mathbb{P}$ and $\mathbb{Q}$ parameters, which need to be estimated accordingly:

$$\Theta^{\mathbb{P}} = \{\kappa_v^{\mathbb{P}}, \kappa_m^{\mathbb{P}}, \theta_m^{\mathbb{P}}, \nu_m^{\mathbb{P}}, \nu_v^{\mathbb{P}}, \mu_Y^{\mathbb{P}}, \sigma_Y^{\mathbb{P}}, \eta_Y\}, \quad \Theta = \{\kappa_v, \kappa_m, \theta_m, \nu_m, \nu_v, \mu_Y, \sigma_Y\}. \tag{12}$$

[5] Pan (2002) argues that introducing different intensities of jumps under the historical and pricing measure introduces a jump-timing risk premium that is very difficult to disentangle from the mean jump risk premium. Our assumption artificially incorporates the jump-timing risk premium into the mean jump size risk premium.

[6] In the literature, σ_Y has sometimes been constrained to be the same under $\mathbb{P}$ and $\mathbb{Q}$ (Bates (1988), Naik and Lee (1990)), but this is not required by the absence of arbitrage (in contrast to σ_v, σ_m, and $\rho_{Y,v}$). We follow Broadie et al. (2007) by allowing them to be different. Indeed, they find strong evidence that their being different has a strong impact on the magnitude of the premium attached to the mean price jump size.

The remaining parameters are, by assumption, equal under both measures:

$$\Theta^{\mathbb{P},\mathbb{Q}} = \{\lambda_0^{Yv}, \lambda_1^{Yv}, \lambda_2^{Yv}, \lambda_0^{m}, \lambda_1^{m}, \sigma_m, \sigma_v, \rho_{Yv}\}. \tag{13}$$

C. Derivatives pricing

Within the class of affine models, option pricing is most efficiently performed using Fourier inversion techniques. As a starting point, we need the characteristic function of the underlying processes. Due to the affine property of the VIX square in Proposition II.1, we have the following result:

Proposition II.2. *In the two-factor stochastic volatility model with jumps defined by equations* (1)-(8)*, the Laplace transforms of the VIX square and the S&P 500 returns are exponential affine in the current values of the factor processes v and m:*

$$\Psi_{VIX_T^2}(t,\tilde{v},\tilde{m};\omega) := \mathbb{E}_t^{\mathbb{Q}}\left[e^{\omega VIX_T^2}\middle| v_t = \tilde{v}, m_t = \tilde{m}\right] = e^{\alpha(T-t)+\beta(T-t)\cdot\tilde{v}+\gamma(T-t)\cdot\tilde{m}},$$
$$\Psi_{Y_T}(t,y,\tilde{v},\tilde{m};\omega) := \mathbb{E}_t^{\mathbb{Q}}\left[e^{\omega Y_T}\middle| y_t = y, v_t = \tilde{v}, m_t = \tilde{m}\right] = e^{\alpha_Y(T-t)+\beta_Y(T-t)\cdot y+\gamma_Y(T-t)\cdot\tilde{v}+\delta_Y(T-t)\cdot\tilde{m}},$$

where the coefficients in the exponentials are functions defined on $[0,T]$ by the ODEs given in Appendix B. The parameter ω belongs to a subset of $\mathbb{C}$ where the above expectations are finite.

Pricing options on the VIX poses technical difficulties that are not encountered when pricing equity options. Given a call option with strike K and maturity T on the VIX at time $t = 0$, we need to calculate

$$C(\text{VIX}_0, K, T) = e^{-rT}\int_0^\infty (\sqrt{v} - K)^+ f_{\text{VIX}_T^2}(v)dv, \tag{14}$$

where $f_{\text{VIX}_T^2}$ is the $\mathbb{Q}$-density of the VIX square at time $t = T$. The square root appearing in the integral as part of the payoff in (14) prevents us from using the Fast Fourier Transform of Carr and Madan (1999). We would need the log of the VIX to be affine, which is incompatible with affine models for log-returns.

However, this problem can be circumvented. Fang and Oosterlee (2008) introduce the Fourier cosine expansion to price index options on the S&P 500. We extend their method to tackle the pricing of VIX options. Our approach to pricing VIX options is comparable to the inversion performed by Sepp (2008a), but it is more parsimonious in the number of computational parameters.

Proposition II.3. *Consider a European-style contingent claim on the VIX index with maturity T and payoff $u_{VIX}(VIX^2) = (\sqrt{VIX^2} - K)^+$. Given an interval $[a_{VIX}, b_{VIX}]$ for the support of the $VIX_T^2|_{v_0,m_0}$ density, the price $P_{VIX}(t_0, VIX_0)$ at time $t = t_0 \geq 0$ of the contingent claim is*

$$P_{VIX}(t_0, VIX_0) = e^{-r(T-t_0)} \sum_{n=0}^{N-1}{}' A_n^{VIX^2} U_n^{VIX^2}, \tag{15}$$

where the prime superscript in the sum $\sum'$ means that the first term $A_0^{VIX^2} U_0^{VIX^2}$ is divided by 2. The terms in the sum are defined by

$$A_n^{VIX^2} = \frac{2}{b_{VIX} - a_{VIX}} Re\left\{ \Psi_{VIX_T^2}\left(t_0, v_0, m_0; \frac{in\pi}{b_{VIX} - a_{VIX}}\right) \exp\left(-ia_{VIX}\frac{n\pi}{b_{VIX} - a_{VIX}}\right)\right\}, \tag{16}$$

$$U_n^{VIX^2} = \int_{a_{VIX}}^{b_{VIX}} u_{VIX}(v) \cos\left(n\pi \frac{v - a_{VIX}}{b_{VIX} - a_{VIX}}\right) dv. \tag{17}$$

The coefficient $A_n^{\text{VIX}^2}$ is computed using Proposition II.2 and $U_n^{\text{VIX}^2}$ is known in closed form and given in Appendix C.

III. Data and preliminary analysis

In this section, we describe our data and point out some important characteristics of VIX options. We also perform a preliminary model selection exercise, in which we try to jointly calibrate some candidate models to the S&P 500 and VIX option markets.

A. Data description

Options on the VIX were introduced in 2006. Our sample period is from March 1, 2006 to October 29, 2010. The option data consist of the daily closing prices of European options on the S&P 500 and VIX, obtained from OptionMetrics. This time series includes both periods of calm and periods of crisis with extreme events. Therefore, it provides an ideal test bed for our candidate models.

Both the S&P 500 and VIX options datasets are treated following the literature, see e.g. Aït-Sahalia and Lo (1998). We only consider options with maturities between one week and one year and delete options quotes that are not traded on a given date. Then, we infer from highly liquid options the futures price using the at-the-money (ATM) put-call parity. This avoids two issues: Making predictions on future dividends, and using futures closing prices which are not synchronized with the option closing prices. Hence, we consider that the underlying of the options is the index futures and not the index itself. We only work with liquid out-of-the-money (OTM) options for the S&P 500 market and only with liquid call options for the VIX market. If the VIX in-the-money (ITM) call is not liquid, we use the put–call parity to infer a liquid VIX ITM call from a more liquid VIX OTM put. Finally, implied volatilities are computed considering futures prices as underlying.[7]

These adjustments leave a total of 383,286 OTM S&P 500 and 43,775 call options on the VIX, with a daily average of 327 S&P 500 options and 37 VIX options. The number of S&P 500 (VIX) options in our dataset on a given date increases with time, with around 170 (5) options at the beginning of the dataset and around 450 (70) options at the end. At the beginning of the sample, there are one or two short maturities (less than six months) available for VIX options and around six maturities for S&P 500 options, with approximately 40 S&P 500 options per maturity slice. At the end of the sample, the VIX options have around five short maturities with a bit more than 10 options trading per maturity. For S&P 500 options, around ten maturities are available per day with around 60 options for one-month maturities and 40 options for the one-year slice. The low number of VIX options compared to the number of S&P 500 options comes from the fact that VIX options only

[7] We remark that VIX option prices do not satisfy no-arbitrage relations with respect to the VIX index, but rather with respect to the VIX futures value. A VIX call option at time t with maturity T is an option on the volatility for the time interval $[T, T + 30d]$, where $30d$ stands for 30 days. The value VIX_t at time t is related to the volatility on the time interval $[t, t + 30d]$, which might not overlap at all with $[T, T + 30d]$.

started trading in 2006. At the end of our sample, the total VIX options volume per day is about one-half the total volume of S&P 500 options traded.

B. Descriptive statistics

Table I presents the first four sample moments of the S&P 500 futures returns and VIX index levels, over two different periods of time. The first period starts in March 2006 and ends in February 2009, i.e., it spans the pre-crisis period as well as the beginning of the crisis. The second period begins in March 2009 and lasts until October 2010. For our estimation, these two periods serve as the in-sample and out-of-sample periods.

The S&P 500 log-returns exhibit a high kurtosis, especially during the in-sample period, suggesting the presence of rare and large movements. In the in-sample period, their skewness is slightly positive, but becomes negative in the out-of-sample period, due to the substantial losses made during the financial crisis. The VIX index exhibits a large positive skewness and kurtosis in the in-sample period. However, in the out-of-sample period, both skewness and kurtosis decrease significantly whereas the mean increases by 45%, indicating that the values of the index are of larger overall magnitude but with less extreme values.

[Table 1 about here.]

Panel A of Figure 1 displays the joint evolution of the S&P 500 and the VIX index from 2006 to 2010. The S&P 500 returns and the VIX daily increments are highly negatively correlated (with a correlation coefficient of -0.832 over this period), which explains the popularity of VIX contracts for hedging part of the equity risk of a portfolio. Panel B represents the expected one-month forward log-returns of the S&P 500 from March 1st, 2006 to October 29th, 2010 as implied by prices of S&P 500 options and calculated following the method of Bakshi et al. (2003). The expected forward returns illustrate the variety of market situations covered by our time series. They were almost constant until the end of 2007, equal to a positive value and thus indicating that market participants were expecting a stable income from investing in the index. From the end of 2007, they exhibit more variation and eventually turn negative. Following the bail-out of Lehman Brothers in September

2008, the expected forward returns drop below -1.5%. Then, they gradually come back and become stable mid-2009 around a slightly negative level close to -0.2%. In 2010, the sudden increase in the VIX index coincides with a further sudden drop of the expected forward returns falling to almost -0.5%.

[Figure 1 about here.]

Even though the S&P 500 and VIX markets are closely related, we emphasize that options on the VIX and S&P 500 differ substantially in their characteristics. First, S&P 500 and VIX derivatives with the same maturity contain information on the S&P 500 over different time periods. While an S&P 500 option with maturity T contains information about the future S&P 500 index level at time T and therefore about the S&P 500 volatility up to T, a VIX option with maturity T embeds information about the VIX at time T and therefore about the S&P 500 volatility between T and $T + 30$ days. Second, the implied volatility smiles backed out from S&P 500 and VIX option prices have different shapes. Panels C and D of Figure 1 display the S&P 500 and VIX smiles on May 10, 2010. The implied volatilities (IVs) are computed using the standard Black–Scholes formula. The VIX IVs are in general substantially higher than the S&P 500 IVs. They range in our sample from 34% to 216% with an average of 80% whereas the S&P 500 IVs range from 6% to 162% with an average of 26%. The implied volatilities of S&P 500 options are in general decreasing with moneyness, which highlights the expensiveness of out-of-the money put options on the S&P 500. As these options provide investors with protection against large downward movements in returns, the negative skewness of the volatility smile reflects their risk aversion towards such movements. Due to the leverage effect, negative changes in returns are strongly related to increases in volatility, which out-of-the-money VIX call options can hedge. This explains why VIX implied volatilities tend to be positively skewed.

The difference between these markets is also reflected by other indicators such as the put–call trading ratio. Almost twice as many puts as calls are traded daily in the S&P 500 options market, but the situation is reversed in the VIX market, where the amount of calls traded daily is almost double that of the puts. In fact, we can observe in Panels C and D of Figure 1 that the log-moneynesses traded for S&P 500 options are mostly negative (which corresponds to OTM put options) and often

positive for VIX options (OTM calls).

C. Joint calibration

Before we bring our models to the time series of data, we carry out a joint calibration exercise using the cross-section of S&P 500 and VIX options on a particular date. This exercise gives us some guidance for model design and allows us to reduce the set of models to be estimated on a time series of options' data. If a model is not flexible enough to jointly reproduce the implied volatility patterns of both markets on a single date, the $\mathbb{Q}$–dynamics of the model is not sufficiently rich to accurately price both the S&P 500 and VIX derivatives jointly, and we can safely discard this model from further consideration.

Let us fix a date t and consider $\{IV^{Mkt}_{\text{SPX},i}\}_{i=1\cdots N_{\text{SPX}}}$, the set of N_{SPX} market implied volatilities of S&P 500 options for strikes $\{K_i\}$ and maturities $\{T_i\}$. We denote by $\{IV^{Mkt}_{\text{VIX},j}\}_{j=1\cdots N_{\text{VIX}}}$ the set of N_{VIX} market implied volatilities of VIX options. To estimate the parameters, we minimize the root mean squared error (RMSE) between the market and model implied volatilities:[8]

$$\text{RMSE}_{\mathcal{M}}(t) := \sqrt{\frac{1}{N_{\mathcal{M}}} \sum_{1\leq i\leq I} \left(IV^{Mkt}_{\mathcal{M},i} - IV^{Mod}_{\mathcal{M},i}\right)^2}, \quad \mathcal{M} \in \{\text{SPX}, \text{VIX}\}, \tag{18}$$

$$\text{RMSE}(t) \quad := \frac{1}{2}\left(\text{RMSE}_{\text{SPX}}(t) + \text{RMSE}_{\text{VIX}}(t)\right). \tag{19}$$

We use two global optimizers to cope with the non-convexity of the calibration problem and the potential existence of multiple local minima, namely the Covariance Matrix Adaptation Evolution Strategy (CMA-ES), introduced by Hansen and Ostermeier (1996), and the Differential Evolution (DE) algorithm introduced by Storn (1996).

For our calibration, we choose a date on which the markets were under stress, namely May 5, 2010 at the beginning of the European sovereign debt crisis. After cleaning our data as described previously, we have 91 VIX options at six different maturities (from 0.04 to 0.46 years) and 486 S&P 500 options

[8] Alternatively, we checked that using distances taking into account the bid–ask spread of IVs as in Cont and Kokholm (2013) does not significantly change the quality of fit. Instead of the RMSE, we also looked at average relative errors (ARE). However, this does not affect our conclusions. The results using ARE are available upon request.

at eleven different maturities (from 0.05 to 0.91 years) available. We emphasize that we perform a joint calibration. Hence, all this data is entered as input to minimize the total RMSE in (19) from the VIX and the S&P 500 market simultaneously across all available maturities and moneyness.

In Figure 2, we plot the market and model implied volatilities for the S&P 500 (Panels A, C, E) and the VIX (Panels B, D, F) for two maturity slices each. For the S&P 500 options, we choose the two maturities $T = 0.05$ and $T = 0.3$, and for the VIX options, $T = 0.04$ and $T = 0.36$. As candidate models, we choose different sub-specifications of the general model (SVJ2) presented in Equations (1)–(8), namely the Heston model and the Heston model with jumps in returns and volatility (SVJ). We find that the jump component in the stochastic central tendency does not improve the results for the SVJ2 model, we therefore present the results with m being a diffusive process.

[Figure 2 about here.]

From Panel A, Figure 2, we observe that the Heston model provides reasonable results for the S&P 500 market. However, for the VIX market (Panel B), the Heston model clearly fails to reproduce one of the stylized facts of VIX option markets: the positive skew of the implied volatility surface. This failure is most pronounced for the short-term options, where the Heston model generates a significant negative skew. The results for the SVJ model look much more promising. Just by adding jump components to the returns and volatility process, we can now generate the positive skew in the VIX market (Panel D), while providing an almost perfect fit for the S&P 500 options market. The SVJ model only struggles at the short end of the VIX implied volatility surface. This shortcoming disappears when we extend the SVJ specification to the SVJ2 model by adding the factor m. Doing so gives us not only a remarkable fit for the S&P 500 but also for the VIX options market (Panel F). Looking at the RMSEs of the SVJ and SVJ2 models, we find that the SVJ provides an RMSE_{SPX} of 1.27% and an RMSE_{VIX} of 11.60%. The SVJ2 model yields 1.17% and 5.15%, respectively. Hence, while the two models are comparable in terms of their performance on the S&P 500 options market, there is an obvious difference in the VIX market on the chosen date.

In unreported results, we perform calibration exercises on other days, also including calm periods. Irrespective of the day, we observe that the SVJ and SVJ2 models perform comparably on the S&P 500 options market, both fitting the data very well. In contrast, we find that there are dates when

the SVJ model struggles to fit the VIX IVs in addition to the S&P 500 IVs whereas the SVJ2 model satisfactorily fits both.[9] Therefore, we conclude from our calibration exercise that we can discard the Heston model from further analysis and that jumps in the volatility are necessary.

IV. Estimation methodology

Daily calibration is essentially a multiple curve fitting exercise to check whether the models can fit the risk-neutral distributions implied by option prices at different maturities. Some of the parameters we get from daily calibrations in the previous section are unstable and vary substantially from one day to the next.[10] To achieve a more robust estimation which is consistent with the whole time series of in-sample data, we choose a methodology based on particle filtering. A particle filter uses a time series of observable market data called measurements , to estimate the conditional densities of unobserved latent processes such as the volatility and jump processes at every point in time during the estimation period. It can be combined with maximum likelihood estimation for parameter estimation and standard error calculations. Using a time series of S&P 500 and VIX indexes and options, we estimate both the $\mathbb{P}$- and $\mathbb{Q}$-dynamics of the model to obtain a set of model parameters that jointly prices spot and options in both markets consistently over time. Before introducing the specific filter used, we specify the discretized model and the measurement errors.

A. Discretized model and specification of errors

We discretize the continuous-time model on a uniform time grid composed of $M+1$ points $t \in \{t_0 = 0, t_1 = \Delta t, ..., t_k = k\Delta t, ..., t_M = M\Delta t\}$, for some $M \in \mathbb{N}^*$. Since we use daily data, Δt corresponds

[9] Our findings are consistent with Gatheral (2008), who shows that the Heston model is incapable of reproducing the positive skew in VIX IVs, and with Sepp (2008a,b), who finds that incorporating positive jumps in the volatility dynamics into the Heston model removes this shortcoming.

[10] Parameters obtained when calibrating to daily options prices are not stable over time, as explained in Broadie et al. (2007) and Lindström et al. (2008).

to one day. In discrete time, the model evolves under $\mathbb{P}$ as follows:[11]

$$\Delta Y_t = [-\lambda^{Yv}(v_t, m_t)(\theta_Z^{\mathbb{P}}(1,0,0)-1) - \frac{1}{2}v_t + \gamma_t]\Delta t + \sqrt{v_t}\Delta W_t^{Y,\mathbb{P}} + Z_t^{Y,\mathbb{P}}\Delta N_t^{Yv}, \tag{20}$$

$$\Delta v_t = \kappa_v^{\mathbb{P}}\left(\frac{\kappa_v}{\kappa_v^{\mathbb{P}}}m_t - v_t\right)\Delta t + \sigma_v\sqrt{v_t}\Delta W_t^{v,\mathbb{P}} + Z_t^{v,\mathbb{P}}\Delta N_t^{Yv}, \tag{21}$$

$$\Delta m_t = \kappa_m^{\mathbb{P}}(\theta_m^{\mathbb{P}} - m_t)\Delta t + \sigma_m\sqrt{m_t}\Delta W_t^{m,\mathbb{P}} + Z_t^{m,\mathbb{P}}\Delta N_t^{m} \tag{22}$$

where the notation ΔX_t for some process X represents the increment $X_{t_{k+1}} - X_{t_k}$ where $t = t_k \in \{t_0, ..., t_{M-1}\}$.

As the log-returns are observable, equation (20) is the first measurement equation. The second measurement equation comes from the observation of the VIX index level with error:

$$\text{VIX}_t^2 - (\alpha_{\text{VIX}^2}v_t + \beta_{\text{VIX}^2}m_t + \gamma_{\text{VIX}^2}) = \epsilon_t^{\text{VIX}}. \tag{23}$$

Jiang and Tian (2007) point to systematic biases in the calculation of the VIX index, such as model misspecification or data limitations. For example, in practice, the index is calculated using a finite number of options thereby inducing an error in the computation of the integral present in the definition. These biases are captured by the error term ϵ_t^{VIX}. It is assumed to follow a normal distribution with mean zero and variance $s > 0$. The last measurements are the prices of S&P 500 and VIX options. We assume that the option prices are observed with an error, which comes from different sources, such as bid–ask spreads, potential model misspecification, timing and processing errors. We define these errors as the relative differences between the market $O_t^{\mathcal{M},Mkt}$ and the model prices $O_t^{\mathcal{M},Mod}$, $\mathcal{M} \in \{\text{SPX}, \text{VIX}\}$:

$$\frac{O_{t,i}^{\text{SPX},Mod}(Y_t, v_t, m_t, \Theta, \Theta^{\mathbb{P},\mathbb{Q}}) - O_{t,i}^{\text{SPX},Mkt}}{O_{t,i}^{\text{SPX},Mkt}} = \epsilon_{t,i}^{\text{SPX},options}, \quad i = 1, \ldots, N_{\text{SPX},t}, \tag{24}$$

$$\frac{O_{t,j}^{\text{VIX},Mod}(v_t, m_t, \Theta, \Theta^{\mathbb{P},\mathbb{Q}}) - O_{t,j}^{\text{VIX},Mkt}}{O_{t,j}^{\text{VIX},Mkt}} = \epsilon_{t,j}^{\text{VIX},options}, \quad j = 1, \ldots, N_{\text{VIX},t}, \tag{25}$$

where $N_{\mathcal{M},t}$ is the number of contracts available in the corresponding market. We assume the error

[11] For the particle filter, we actually use a Milstein scheme to improve the precision of the discretized dynamics. See Kloeden and Platen (1992) for details.

terms to be normally distributed and heteroscedastic:

$$\epsilon_{t,i}^{\text{SPX},options} \sim \mathcal{N}(0, \sigma^2_{\epsilon_{t,i}^{\text{SPX}}}), \quad \epsilon_{t,j}^{\text{VIX},options} \sim \mathcal{N}(\mu_{\epsilon_t^{\text{VIX}}}, \sigma^2_{\epsilon_{t,j}^{\text{VIX}}}), \tag{26}$$

where $\mu_{\epsilon_t^{VIX}}$ is proportional to the error ϵ_t^{VIX} which has been made on the estimation of the VIX level. Indeed, if the underlying's value is not accurately estimated, it introduces a bias in the valuation of VIX options. The variance of the errors is

$$\sigma^2_{\epsilon_{t,i}^{\text{SPX}}} = \exp\left(\phi_0 \cdot \text{bid-ask spread}_i + \phi_1 \left|\log\left(\frac{K_i}{F_t^{\text{SPX}}(T_i)}\right)\right| + \phi_2(T_i - t) + \phi_3\right), \tag{27}$$

$$\sigma^2_{\epsilon_{t,j}^{\text{VIX}}} = \exp\left(\psi_0 \cdot \text{bid-ask spread}_j + \psi_1 \left|\log\left(\frac{K_j}{F_t^{\text{VIX}}(T_j)}\right)\right| + \psi_2(T_j - t) + \psi_3\right), \tag{28}$$

with ϕ_i and ψ_i in $\mathbb{R}$, $i \in \{0, ..., 3\}$.[12]

B. Particle filter

At every period $t = t_n$, the measurement vector y_t collects observed market prices. By $y^t = (y_{t_0}, ..., y_{t_n})$, we denote all the observations available up to time t. The filtering problem consists of recursively approximating the distribution of the latent state L_t,

$$L_t = \left\{v_t, m_t, \Delta N_t^{Yv}, \Delta N_t^m, Z_t^{Y,\mathbb{P}}, Z_t^{v,\mathbb{P}}, Z_t^{m,\mathbb{P}}\right\}, \tag{29}$$

conditional on y^t. Particle filters are perfectly adapted to our problem: They can handle observations that are nonlinear functions of latent variables as well as equations with non-Gaussian innovations.

There are many types of particle filters. We use the Auxiliary Particle Filter (APF) proposed by Pitt and Shephard (1999). Compared to more basic particle filters, such as the Sampling Importance Resampling (SIR) filter, the APF is better suited to detect jumps, whereas the SIR filter faces sample impoverishment leading to potential particle degeneracy. Both filters are described in Johannes et al. (2009) for filtering latent factors from returns in a Heston model with jumps in returns.

[12]The fact that the option pricing errors are normally distributed does not constitute a restriction. The reason is that the errors are heteroscedastic and the coefficients generating the heteroskedasticity are driven by the data, i.e., we optimize over the parameters $\{\phi_i, \psi_i\}_{0 \leq i \leq 3}$.

We develop an extension of their algorithm that is able to handle more data (the VIX market data on top of the S&P 500 market data) as well as the second volatility factor m and the volatility jumps. The likelihood estimation and particle filter are presented in detail in Appendix D. Furthermore, we performed additional data treatments for S&P 500 and VIX options before running the particle filter. They are described in Appendix E.

C. Candidate models and datasets

For the convenience of illustrating the effect of the incremental information contained in the different markets, we proceed with our empirical investigation and develop insights in a pedagogical way by defining four different datasets as follows:

	S&P 500 index	S&P 500 index and options
VIX index	$\mathcal{D}_1$	$\mathcal{D}_2$
VIX index and options	$\mathcal{D}_3$	$\mathcal{D}_4$

$\mathcal{D}_1$ only contains data on the S&P 500 and VIX indexes. $\mathcal{D}_2$ (resp. $\mathcal{D}_3$) also contains S&P 500 options (resp. VIX options). Finally, $\mathcal{D}_4$ contains all available data. Splitting up the data this way allows us to draw inferences on the information contents of the different markets and to study whether these contents are consistent with one another.

As the calibration exercise described in Section III.C has shown, the SVJ and SVJ2 model perform well in jointly fitting the S&P 500 and VIX market on a particular day. Therefore, these two models are natural candidates to analyze with the particle filter. In addition, to appreciate the impact of jumps, we consider a two-factor volatility model that has no jumps. We label this model SV2.[13]

[13] In unreported results, we do not find evidence that the inclusion of additional jumps in the central tendency factor helps to capture the time series of option prices. Therefore, as in the calibration exercise of the previous section, we restrict our analysis of the SVJ2 model by setting $J_t^m = 0$ for all t.

V. Results and discussion

After presenting the parameter estimates, we analyze the trajectories of the latent factors, depending on the model and dataset considered. We also investigate the benefits of including the jumps and stochastic central tendency by analyzing the in and out-of-sample option pricing errors. We then discuss whether the VIX market and S&P 500 market convey the same information about the volatility dynamics. Finally, we study the dynamics of equity and variance risk premia.

A. Likelihood criteria and parameter estimates

Tables II and III present the point estimates and standard errors resulting from the estimation of the SVJ2 model and its SVJ and SV2 sub-specifications to datasets $\mathcal{D}_1$ to $\mathcal{D}_4$. The last rows of Table III indicate the log-likelihood values and the values of the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) for each estimation.

Before we discuss specific parameters, we make the following remark. When we estimate the models with index data ($\mathcal{D}_1$) only, the likelihood criteria are slightly in favor of the SVJ2 model. However, the $\mathbb{Q}$-parameters and $\kappa_v^{\mathbb{P}}$ are difficult to estimate. Their standard errors are typically four to five times larger than those obtained with the other datasets. Therefore, we conclude that $\mathcal{D}_1$ is not rich enough to provide a reliable estimation of the models considered. In particular, despite the fact that the VIX index is constructed from option prices, it does not contain enough information to accurately infer the parameters which characterize the $\mathbb{Q}$-distribution of the S&P 500 returns. Consequently, $\mathcal{D}_1$ is not informative enough to provide reasonable pricing performance for S&P 500 options.

[Table 2 about here.]

[Table 3 about here.]

Given this result for $\mathcal{D}_1$, we focus our discussion on the results for the larger datasets $\mathcal{D}_2$ to $\mathcal{D}_4$. We start by analyzing the parameters that we assume to be equal under both probability measures. The estimates for the jump intensities suggest that the dominant factor driving the intensities is m. Indeed, the estimates of λ_2^{Yv} range between 3.64 and 4.25 for the SVJ2 model and are significantly

different from zero. In contrast, the estimates of λ_1^{Yv} range between 0.15 and 1.1 and are not significant for any dataset. For the SVJ model with one volatility factor, the effect of m is transferred to v and we can expect an increase in the estimates for λ_1^{Yv}. Indeed, they range between 2.38 and 2.85 and are significant. When we use the whole dataset $\mathcal{D}_4$, the constant term λ_0^{Yv} is significantly different from zero for both the SVJ2 and SVJ model.

For the second volatility factor m, we find that the volatility parameter σ_m is in the interval $[19\%, 35\%]$ for both the SV2 and SVJ2 models regardless of the dataset chosen for the estimation. This value is about one-half of the volatility factor σ_v found for the process v in these two models. In addition, the speed of mean-reversion of the process m is about twenty times smaller than that of v under both measures. Hence, we can interpret the process v as a factor representing erratic short-term fluctuations of the variance, whereas the process m is more persistent and captures smoother medium- to long-term trends.[14]

In the SVJ model, the volatility-of-volatility parameter σ_v tends to range between the estimates σ_m and σ_v of the two-factor models.[15] Not surprisingly, we also find a prominent leverage coefficient ρ_{Yv} across all models and datasets.

In line with the literature, the equity risk premium coefficient η_Y is positive across all models and datasets, which implies a positive diffusive equity risk premium. We find that under $\mathbb{P}$, the mean jump size of returns $\mu_Y^{\mathbb{P}}$ and its variance $\sigma_Y^{\mathbb{P}}$ are difficult to identify, because the likelihood is not very sensitive to a change in their value. Hence, we have decided to estimate them using high-frequency returns. Five-minute returns were obtained from the TAQ database and intra-daily jump occurrences

[14] Indeed, under the assumption that jumps have a minor impact on this expectation compared to the drift term, i.e. $\kappa_v^{\mathbb{P}} >> \lambda_1^{Yv}\nu_v^{\mathbb{P}}, \kappa_v^{\mathbb{P}} >> \lambda_2^{Yv}\nu_v^{\mathbb{P}}$ and $\kappa_v\theta_m^{\mathbb{P}} >> \lambda_0^{Yv}\nu_v^{\mathbb{P}}$ (inequalities satisfied by our parameter estimates), the conditional expectation of the variance $\mathbb{E}_t^{\mathbb{P}}[v_T]$ can be written as:

$$\mathbb{E}_t^{\mathbb{P}}[v_T] \approx \theta_m^{\mathbb{P}}\frac{\kappa_v}{\kappa_v^{\mathbb{P}}} + c\cdot e^{-\kappa_m^{\mathbb{P}}(T-t)}(m_t - \theta_m^{\mathbb{P}}) + e^{-\kappa_v^{\mathbb{P}}(T-t)}(v_t - c\cdot m_t + c\cdot\frac{\kappa_m^{\mathbb{P}}}{\kappa_v^{\mathbb{P}}}\theta_m^{\mathbb{P}}),$$

for a constant $c = \frac{\kappa_v}{\kappa_v^{\mathbb{P}} - \kappa_m^{\mathbb{P}}}$. As $\kappa_v^{\mathbb{P}} >> \kappa_m^{\mathbb{P}}$, the coefficients in front of v_t decays much faster than the one in front of m_t. For $T-t$ equal to three months, $e^{-\kappa_v^{\mathbb{P}}(T-t)}$ is around 0.15, but goes down to 0.03 for six months, and is of order of magnitude of 10^{-3} for a year. In contrast, $e^{-\kappa_m^{\mathbb{P}}(T-t)}$ is around 0.85 for $T-t$ equal to six months, and as high as 0.77 for one year. Therefore the deviation of m_t relative to its long-term mean drives the medium- to long-term expectation of the variance.

[15] We impose the Feller condition $2\kappa_v m_0 > \sigma_v^2$ on the SVJ model, where m_0 is the level of reversion of the variance when the central tendency is a constant process. As a consequence, for all datasets containing options, the estimated volatility-of-volatility parameter σ_v is considerably smaller for the SVJ than for the SV2 and SVJ2.

and sizes were estimated following the method described in Bollerslev and Todorov (2011). Daily jumps were obtained by aggregation of the high-frequency jumps. The resulting moments of the jumps are: $\mu_Y^{\mathbb{P}} = 0.0002$ and $\sigma_Y^{\mathbb{P}} = 0.0039$. Their values were fixed throughout the filtering exercise. Jump size estimates under $\mathbb{Q}$ are statistically significant and strongly negative (around -10%), which is due to investors' risk aversion to jumps, and yields a non-zero jump risk premium. Similarly, the volatility of the returns' jump sizes is larger under $\mathbb{Q}$ than under $\mathbb{P}$. This finding for jumps in returns seems to contrast with the result for the jumps in volatility. There is almost no difference between the estimates of $\nu_v^{\mathbb{P}}$ and ν_v. For the SVJ2 model with the full dataset $\mathcal{D}_4$, we find a mean jump of 0.03 under both measures. However, the estimate for ν_v is highly significant, while the estimate for $\nu_v^{\mathbb{P}}$ is not statistically different from zero.

The diffusive part of the volatility risk premium $\eta_v = \kappa_v - \kappa_v^{\mathbb{P}}$ is mostly negative. Its amplitude, however, depends on the model used. In particular, we find much smaller magnitudes with the SVJ2 and SV2 models than with the SVJ model. The estimates for dataset $\mathcal{D}_1$ illustrate, again, the problem inherent with an estimation using indexes only. The estimates for the variance risk premium η_v vary considerably from strongly negative to strongly positive for different models. For the diffusive part of the stochastic central tendency risk premium $\eta_m = \kappa_m - \kappa_m^{\mathbb{P}}$, we find that its sign changes with the model used, and that it is never statistically significantly away from zero, due to the rather large standard errors of the κ_m and $\kappa_m^{\mathbb{P}}$ parameters.

Finally, the last rows of Table III present the log-likelihood values for the different models and datasets. All likelihood criteria imply that the SVJ2 model significantly outperforms the SVJ and SV2 models when options are included in the estimation dataset. Furthermore, the two-factor models (SVJ2 and SV2) are substantially superior to the SVJ model in fitting the S&P 500 options.

B. Filtered trajectories

Figure 3 displays the filtered trajectories of the volatility process resulting from estimating all three models SVJ, SV2, and SVJ2 using $\mathcal{D}_4$. Whether we focus on the in-sample or out-of-sample period, we find that the filtered volatility values $\sqrt{v_t}$ are consistent across model specification.

[Figure 3 about here.]

We also compare the filtered trajectories of the volatility when using different estimation datasets. By doing so, we can identify the information content of each dataset regarding the variance of S&P 500 returns. In Panels B to D of Figure 3, we plot the difference between the filtered variance using $\mathcal{D}_4$ and the one filtered using $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$, under the SVJ2 specification. Up until the peak in the VIX towards the end of the in-sample period, the difference between the filtered variances is small (less than 3%). During this period, the filtered variance using $\mathcal{D}_1$, which uses only the S&P 500 and VIX index returns, is in general slightly smaller than the variance filtered using the other datasets (Panel B). Hence, the inclusion of options tends to moderately increase the filtered variance. With the start of the financial crisis in the fall of 2008, some new patterns emerge. While the variance filtered using $\mathcal{D}_4$ remains close to the one filtered using $\mathcal{D}_2$ (Panel C), the variance filtered using $\mathcal{D}_1$ and $\mathcal{D}_3$ (Panels B and D) is substantially larger (up to 25 percentage points). In the out-of-sample period and after the variance crisis peak, the difference between the trajectories remains within an interval of ± 3 percentage points except during the period surrounding the second variance increase in May 2010. Hence, in times of market distress, S&P 500 options contain information on the variance process that is different from the one contained in $\mathcal{D}_1$, and more interestingly in $\mathcal{D}_3$ (i.e., VIX options). This observation has statistical significance. Indeed, during those times, the lower bound of the 95% confidence interval of the variance estimated with $\mathcal{D}_3$ is more than 5% higher the upper bound of the variance estimated with $\mathcal{D}_2$. This phenomenon only occurs during the two variance peaks in our sample period. Possible explanations include a difference in the constraints or information sets of the traders involved, which may affect the prices of options on one market differently from those in the other market.

In Figure 4, Panels A and B display the time series of the filtered probability of jumps in the variance process $\{\mathbb{P}(\Delta N_{t_k}^{Y,v} = 1)\}_{k=0,\ldots,M}$, inferred from $\mathcal{D}_1$ and $\mathcal{D}_4$.[16] If this probability is above 10%, we plot the corresponding jump size in Panels C and D. As for the contribution of jumps to the variation of the variance, Figure 4 shows that a substantial number of jumps are filtered during the crisis regardless of the dataset used for estimation. When options are part of the dataset, the largest

[16] We do not report the results for $\mathcal{D}_2$ and $\mathcal{D}_3$, as the results are very similar to $\mathcal{D}_4$.

variance jumps are between 5% and 10%.

[Figure 4 about here.]

Although the volatility trajectories are similar across models (Panel A, Figure 3), the SVJ model generates more volatility jumps during the crisis. Given that the SVJ model has a constant level of reversion for the volatility, jumps are necessary to generate large volatility increases during the crisis. This is not the case for the SV2 (and SVJ2) models that can (also) generate increases in volatility by having a high level of reversion m "pulling" the volatility up. Furthermore, we impose the Feller condition on the SVJ model, which restricts the amplitude of the diffusive volatility movements.

In Figure 5, we plot the trajectories of the central tendency process m using different datasets. Interestingly, the estimation of m is more sensitive to the choice of datasets. Moreover, in contrast to the estimation of v, we observe significant differences depending on which model we take. The SV2 model indeed tends to generate higher values for the central tendency, in particular in 2009 just after the peak in the process v. Intuitively, especially during volatile times, the SV2 model needs to compensate its inability to generate jumps by increasing m_t.

Irrespective of the dataset, we observe that the process m is overall more persistent than the process v, in line with the parameter estimates found. v increases dramatically during the crisis (from September 2008 on) but gradually returns to pre-crisis levels at the beginning of 2010. In contrast, the central tendency starts to increase at the beginning of 2009, a few months after the spike observed in v. On the other hand, the process m does not return to its pre-crisis level of 1% to 2%, but remains between 5% and 8% until the end of the time series.

[Figure 5 about here.]

Even though the central tendency process is more persistent than v, approximating m by a constant as in the SVJ model is too rough of an approximation. In particular, Figure 5 shows that the levels reached by the central tendency during and after the crisis are substantially underestimated by the SVJ model. In fact, the constant central tendency estimated in the SVJ model seems to be close to the average filtered central tendency for the SV2 and SVJ2 models over the in-sample period. Therefore the SVJ model is insensitive and unadaptable to different long-term volatility regimes.

C. *Pricing errors*

We analyze pricing errors, both across model specifications and datasets. Our main focus, however, is on the simultaneous pricing performance of different models for options on both the S&P 500 and VIX.[17]

We first investigate how the different models reproduce S&P 500 option prices over time, looking at different moneynesses and maturities. Focusing first on dataset $\mathcal{D}_2$ in Table IV, we find that the SVJ2 and SV2 are superior to the SVJ. Indeed, the SVJ model exhibits higher Root Mean Square Relative Errors (RMSREs) than the SV2 and SVJ2 models for all option categories except ATM options.[18] The SVJ prices fairly well short-term options (still not as well as the SVJ2 and SV2) but struggles to accurately price deep OTM calls and long-maturity options. Hence, introducing a stochastic central tendency allows us to price long-term and deep OTM S&P 500 options more accurately. This finding supports the interpretation that the process m captures the long-term trends of volatility and therefore helps to better reproduce the term structure of S&P 500 option prices. The SVJ2 and SV2 have similar in-sample pricing errors overall. The SV2 is slightly better at pricing ATM and deep OTM calls, but the SVJ2 prices OTM puts more accurately. Out-of-sample, the SVJ2 model outperforms the SV2 model in most categories of moneyness and maturity. Exceptions are deep OTM puts and long-term options. However, for those categories, the differences in RMSRE do not exceed 2%. Therefore, the SVJ2 model does not overfit the data in the in-sample-period, which would otherwise translate into a deterioration of its performance in the out-of-sample performance compared to the other models.

[Table 4 about here.]

Focusing on the pricing errors of VIX options using the dataset $\mathcal{D}_3$, we find that the SVJ2 model again outperforms all models in terms of RMSRE for most option categories. In-sample, the SVJ2

[17] We also analyzed model generated values of the VIX index. We find that all three models, SVJ, SV2, and SVJ2, accurately reproduce the time series of the VIX index irrespective of which estimation dataset is used. Hence, jumps and a stochastic central tendency appear superfluous for reproducing the trajectory of the VIX level, a result which is confirmed by a Diebold–Mariano test. Detailed results are available upon request.

[18] As the model has been estimated to relative errors, it is sensible to use the same measure of error to evaluate its performance. We found that an assessment in terms of Root Mean Square Errors may be misleading as it focuses on expensive options, i.e., options which are closer to the ATM level.

always performs better than the SVJ model. The SV2 model is only slightly better for deep OTM call options. Interestingly, the SVJ does better than the SV2 at pricing deep OTM VIX calls, indicating that jumps are essential to represent accurately the tail of the volatility distribution. However, the central tendency factor significantly improves the pricing of all other moneyness levels. These observations are confirmed out-of-sample. The addition of a central tendency factor improves the pricing of VIX options for all moneyness levels except for deep OTM calls. In fact, for this category, the SVJ model even performs better than the SVJ2 out-of-sample. Moreover, and consistent with the results obtained when using $\mathcal{D}_2$ as estimation dataset, the SV2 model substantially outperforms the SVJ model in pricing VIX options with a maturity exceeding two months.

When using all the available data $\mathcal{D}_4$, the SVJ2 model yields significantly smaller in-sample RMSREs than the SVJ and SV2 models, for most categories of S&P 500 and VIX options. We find that the two-factor models perform much better than the SVJ to price deep OTM puts and calls on the S&P 500 as well as long maturity options. In turn, the SVJ model outperforms the SV2 model in fitting deep OTM calls on the VIX. While the SVJ2 outperforms the SV2 model at pricing S&P 500 options in- and out-of-sample, it appears that the SV2 model slightly outperforms the SVJ2 model in fitting out-of-sample VIX options.

To test whether the pricing performance is, on average, significantly better for the SVJ2 than for its nested models, we use the Diebold and Mariano (1995) test (DM).[19] We consider two loss functions: the Mean Square Error (MSE) of the option price errors, and their Mean Square Relative Error (MSRE).

For S&P 500 options and using dataset $\mathcal{D}_2$, the results in Panel A of Table V confirm that the SVJ2 model provides significantly better in-sample and out-of-sample MSEs than the other two models. The DM tests for the MSRE loss function are not as significant, but still positive and larger out-of-sample than in-sample. When we switch to the dataset $\mathcal{D}_4$, the results confirm that the SVJ2 model has smaller in-sample pricing errors than the other two models, especially for S&P 500 options. However, the test indicates that the SVJ2 model does not outperform the SV2 model out-of-sample,

[19]The DM test works as follows. Consider a loss function, e.g., $L(e_t) = |e_t|$, where e_t is the difference at time t between two model pricing errors. If these two models have comparable pricing errors, then the expectation of their loss differential should be zero. The DM test provides a test statistic for this differential.

due to the slight lack of precision in pricing OTM and deep OTM calls.

[Table 5 about here.]

For VIX options (and using $\mathcal{D}_4$), the results in Panel B of Table V do not favor one model over the others when considering the MSE as distance criterium. In terms of MRSE, they are slightly in favor of using the SVJ2 model, as it better fits the cheap OTM options.

We can draw two main conclusions from our analysis of the pricing errors. First, including a stochastic central tendency in the model significantly improves the pricing of long-term options and the representation of the tails of the distributions of the returns (OTM puts and calls on the S&P 500). Second, jumps improve the representation of the right tail of the variance distribution (OTM calls on the VIX) as well as of the short-term options.

Given that we use more than four and a half years of option data on two different markets with a wide range of moneynesses, the pricing errors resulting from our time series estimation are relatively small. However, our analysis also uncovers some potential shortcomings of the affine framework. Even the SVJ2 model has difficulty reproducing the observed volatility smiles during and after the crisis. In particular, OTM puts on the S&P 500 tend to be underpriced and OTM calls are generally overpriced, the model-implied volatility smile does not have enough skewness. This phenomenon affects short-maturity options in particular, and indicates that the model struggles to reproduce higher moments at the short end.

To further illustrate this point, Figure 6 compares the skewness and kurtosis of the S&P 500 returns implied by the market and model option prices, the models being calibrated to the full dataset $\mathcal{D}_4$. While the skewness of the returns is well represented at the beginning of the in-sample period, it is underestimated from late 2007 until the end of our sample. In the out-of-sample period, this phenomenon becomes much more apparent, as all three models yield an implied skewness which is about one-half of that implied by the market. The SVJ2 model provides a slight improvement over the other two models, but is still far from reality. Similarly, the kurtosis is only slightly underestimated at the beginning of the time series, but in the out-of-sample period the model kurtosis is about one-half of that implied by the market. Additionally, we find that having VIX options in the estimation

dataset does not improve the representation of the S&P 500 implied moments.

[Figure 6 about here.]

D. Market integration

Even for affine models, particle filter based estimation of option pricing models is computationally intensive. Therefore, the recent literature has proposed estimating option pricing models by only using the S&P 500 and VIX index data, thereby avoiding the computational burden associated with option valuation. See, e.g., Duan and Yeh (2010, 2011). However, as we mentioned in Section V.A, the estimation results indicate that using index data alone may lead to erroneous estimated dynamics, and, consequently, erroneous option prices. Indeed, Table IV reports RMSREs for S&P 500 options, which are two to five times larger when estimating from $\mathcal{D}_1$ than when estimating from $\mathcal{D}_2$. For VIX options, the RMSREs are two to four times larger when using $\mathcal{D}_1$ than when using $\mathcal{D}_3$. This mispricing is particularly marked for the SVJ2 and SV2 models, which have more parameters and require more information for their estimation. Hence, we conclude that one should refrain from estimating an option pricing model using only the S&P 500 and VIX index data.

Although the VIX and S&P 500 options provide different information on the trajectory of volatility in times of market turmoil (Figure 3, Panels C and D), jumps filtered from estimating the SVJ2 with $\mathcal{D}_2$ and $\mathcal{D}_3$ present similar patterns, with slightly less jumps filtered from S&P 500 options. In the fall of 2008, several jumps around 10% occur with a probability larger than 10%. Finally, slightly smaller jumps are detected when the Eurozone sovereign debt crisis emerges in May 2010.

Despite these similarities, Table IV shows that the RMSREs on S&P 500 options are four to five times higher when estimating the two-factor models using $\mathcal{D}_3$ than when estimating them using $\mathcal{D}_4$. This ratio is much smaller for the SVJ model, partly because its estimation with $\mathcal{D}_2$ already yields rather bad results. Therefore, our results suggest that VIX options do not contain enough information to price S&P 500 options accurately. Estimating the SVJ2 and SV2 models using $\mathcal{D}_3$ in fact leads to severely mispriced OTM S&P 500 calls, which indicates that VIX options contain very little information on the right tail of the distribution of the S&P 500 returns. Concerning deep

OTM puts on the S&P 500, it is striking to see that the estimation using $\mathcal{D}_3$ outperforms the one using $\mathcal{D}_2$ out-of-sample. This indicates that VIX options provide valuable information on the left tail of the returns' distribution. Conversely, we note that the RMSREs on VIX options using $\mathcal{D}_2$ are about 50% higher than those using $\mathcal{D}_3$, both in-sample and out-of-sample. Therefore, we conclude that S&P 500 options do not include all the information contained in VIX options.[20]

E. Variance risk premium

Following Bollerslev and Todorov (2011), we define the annualized integrated variance risk premium (IVRP) as

$$\text{IVRP}(t,T) = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(QV_{[t,T]}\right) - \mathbb{E}_t^{\mathbb{Q}}\left(QV_{[t,T]}\right)\right],$$

where $QV_{[t,T]}$ denotes the quadratic variation of the log price process, which is the sum of the integrated variance of the returns and the squared jumps in the time interval considered:

$$QV_{[t,T]} = \int_t^T v_s ds + \sum_{i=N_t^{Yv}}^{N_T^{Yv}} \left(Z_i^Y\right)^2$$

The IVRP represents the expected payoff when buying a variance swap at time t with maturity T. We can further decompose the IVRP into a continuous and a discontinuous part:

$$\text{IVRP}(t,T) = \text{IVRP}^c(t,T) + \text{IVRP}^d(t,T)$$

with

$$\begin{aligned}
\text{IVRP}^c(t,T) &= \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(\int_t^T v_{s^-} ds\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\int_t^T v_{s^-} ds\right)\right], \\
\text{IVRP}^d(t,T) &= \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(\sum_{i=N_t^{Yv}}^{N_T^{Yv}} \left(Z_i^Y\right)^2\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\sum_{i=N_t^{Yv}}^{N_T^{Yv}} \left(Z_i^Y\right)^2\right)\right].
\end{aligned}$$

[20] This result does not seem to be model-driven. Indeed, the RMSREs on VIX options are approximately all reduced when estimating models from $\mathcal{D}_2$ to $\mathcal{D}_4$ while the RMSREs on SPX options remain more or less the same. Therefore, the models are better identified when using $\mathcal{D}_4$ than with $\mathcal{D}_2$.

Each part can also be decomposed into a contribution from m_t and another one from v_t:

$$\begin{aligned}
\text{IVRP}^c(t,T) &= \left((A^{\mathbb{P}}-A)v_t + (B^{\mathbb{P}}-B)m_t + G^{\mathbb{P}} - G\right), \\
\text{IVRP}^d(t,T) &= \left(\lambda_1^{Yv}A^{\mathbb{P}}v_t + (\lambda_1^{Yv}B^{\mathbb{P}} + \lambda_2^{Yv}\hat{A}^{\mathbb{P}})m_t + \lambda_0^{Yv} + \lambda_1^{Yv}G^{\mathbb{P}} + \lambda_2^{Yv}\hat{B}^{\mathbb{P}}\right)\left[(\sigma_Y^{\mathbb{P}})^2 + (\mu_Y^{\mathbb{P}})^2\right] \\
&\quad - \left(\lambda_1^{Yv}Av_t + (\lambda_1^{Yv}B + \lambda_2^{Yv}\hat{A})m_t + \lambda_0^{Yv} + \lambda_1^{Yv}G + \lambda_2^{Yv}\hat{B}\right)\left[(\sigma_Y)^2 + (\mu_Y)^2\right].
\end{aligned}$$

The coefficients under $\mathbb{Q}$ are given in Appendix A.[21]

In Figure 7, Panel A, we plot the evolution of the IVRP generated by the SVJ2 model using our largest dataset $\mathcal{D}_4$. In line with the literature, we find that the IVRP is strongly negative with a sharp drop at the end of 2008.[22] Afterwards, the IVRP steadily shrinks but remains larger - in absolute terms - than before the drop. In Panel B of Figure 7, we decompose the instantaneous variance risk premium into its continuous and discontinuous parts. The discontinuous component of the IVRP dominates for shorter maturities, indicating that including jumps helps to represent the shorter end of the variance premium's term structure. At the long end, the effect of a positive jump in the variance process is likely to be dampened by the reversion of the variance to its long-term mean.

[Figure 7 about here.]

Finally, we find that the contribution of the stochastic central tendency to the continuous part of the IVRP is negligible for short maturities but plays a substantial role for maturities larger than three months (see Panel C of Figure 7). The central tendency m usually sets the level of the continuous IVRP, but it becomes secondary when the market conditions change and the variance peaks . The central tendency also plays a major role in the discontinuous part of the IVRP and determines most of its level especially during quiet times, see Panel D of Figure 7. Hence, both the variance jumps and the stochastic central tendency play a crucial role in the IVRP. While the jumps help to represent the short end of the variance term structure, it is mostly the stochastic central tendency m that determines the IVRP during calm market periods. In times of financial crises, the process v

[21] The time-to-maturity τ_{VIX} needs to be replaced by $(T-t)$. The expressions remain similar under $\mathbb{P}$. $\mathbb{Q}$-parameters simply need to be replaced by their $\mathbb{P}$-counterpart, except $B^{\mathbb{P}}$ and $G^{\mathbb{P}}$ in which κ_v remains under $\mathbb{Q}$.

[22] This result holds across datasets and models. However, we only plot the results for the SVJ2 using $\mathcal{D}_4$.

contributes the most to setting the IVRP.

VI. Conclusion

In this paper, we carried out an extensive empirical investigation of the information contained in related derivatives markets on the dynamics of volatility. We estimated various specifications of a flexible affine model using a time series of S&P 500 and VIX indexes as well as option prices on both indexes. To extract as much information about extreme events as possible, we used S&P 500 and VIX options with a unique wide range of moneynesses. Instead of a step-wise estimation, we departed from most of the literature and estimated the historical and the risk-neutral parameters jointly, in a single step. Our maximum likelihood estimation procedure is based on particle filtering.

We show that using a model with a stochastic central tendency and jumps in the returns and volatility provides significant improvements for pricing S&P 500 and VIX options jointly, both in- and out-of-sample. Adding a stochastic central tendency helps to better represent the tails of the returns' distributions, as well as the term structure of the S&P 500 and VIX option prices, while jumps allow more flexibility to match the right tail of the variance distribution as well as short-dated options. Regarding the variance risk premia (VRP), jumps additionally play an important role and drive the short term VRP. For longer term horizons, the continuous part of the VRP dominates and the contribution of m is largest when markets are calm.

We investigate the ability of information from one market to explain prices observed in other related markets. We find that the VIX index does not provide an accurate representation of the information contained in the S&P 500 options. Indeed, an estimation using only the VIX levels in addition to S&P 500 returns (hence excluding options) does not allow to reproduce either the S&P 500 or VIX option prices. Similarly, the information in S&P 500 derivatives does not span the information contained in VIX derivatives, and the same holds the other way around. This lack of integration is particularly prominent during market stress, where it is therefore crucial to include the underlyings as well as the derivatives on both markets in order to estimate a model and account for the cross-section of instruments.

Finally, we find that even the model with a stochastic central tendency and jumps is not able to fully reproduce the skewness and kurtosis of the underlying S&P 500 index in times of market turmoil. We conjecture that this limitation is due to the affine nature of our modeling framework. However, departing from this framework generates tremendous additional computational complexity for the particle filter estimation. We leave this challenging avenue for future research.

References

Aït-Sahalia, Yacine, Mustafa Karaman, and Loriano Mancini, 2012, The term structure of variance swaps, risk premia and the expectation hypothesis, Technical report, Princeton University and Swiss Finance Institute.

Aït-Sahalia, Yacine, and Robert Kimmel, 2007, Maximum likelihood estimation of stochastic volatility models, *Journal of Financial Economics* 83, 413–452.

Aït-Sahalia, Yacine, and Andrew Lo, 1998, Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices, *Journal of Finance* 53, 499–547.

Alizadeh, Sassan, Michael W. Brandt, and Francis X. Diebold, 2002, Range-based estimation of stochastic volatility models, *Journal of Finance* 57, 1047–1091.

Andersen, Torben G., Luca Benzoni, and Jesper Lund, 2002, An empirical investigation of continuous-time equity return models, *Journal of Finance* 57, 1239–1284.

Andrews, Donald, 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, Stock return characteristics, skew laws, and the differential pricing of individual equity options, *Review of Financial Studies* 16, 101–143.

Bates, David S., 1996, Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options, *Review of Financial Studies* 9, 69–107.

Bates, David S., 2000, Post-87 crash fears in the S&P 500 futures options, *Journal of Econometrics* 94, 181–238.

Bates, David S., 2012, U.S. stock market crash risk, 1926-2010, *Journal of Financial Economics* 105, 229 – 259.

Bayer, Christian, Jim Gatheral, and Morten Karlsmark, 2013, Fast Ninomiya-Victoir calibration of the double-mean-reverting model, *Quantitative Finance* 13, 1813–1829.

Bollerslev, Tim, and Viktor Todorov, 2011, Tails, fears and risk premia, *Journal of Finance* 66, 2165–2211.

Breeden, Douglas T., and Robert H. Litzenberger, 1978, Prices of state-contingent claims implicit in option prices, *Journal of Business* 51, 621–651.

Broadie, Mark, Mikhail Chernov, and Michael Johannes, 2007, Model specification and risk premia: Evidence from futures options, *Journal of Finance* 62, 1453–1490.

Carr, Peter, and Dilip B Madan, 1999, Option valuation using the fast fourier transform, *Journal of Computational Finance* 2, 1–18.

Chernov, Mikhail, A. Ronald Gallant, Eric Ghysels, and George T. Tauchen, 2003, Alternative models of stock price dynamics, *Journal of Econometrics* 116, 225 – 257.

Christoffersen, Peter, Steven Heston, and Kris Jacobs, 2009, The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well, *Management Science* 55, 1914–1932.

Christoffersen, Peter, Kris Jacobs, and Karim Mimouni, 2010, Models for S&P 500 dynamics: Evidence from realized volatility, daily returns and options prices, *Review of Financial Studies* 23, 3141–3189.

Chung, San-Lin, Wei-Che Tsai, Yaw-Huei Wang, and Pei-Shih Weng, 2011, The information content of the S&P 500 index and VIX options on the dynamics of the S&P 500 index, *The Journal of Futures Markets* 31, 1170–1201.

Cont, Rama, and Thomas Kokholm, 2013, A consistent pricing model for index options and volatility derivatives, *Mathematical Finance* 23, 248–274.

Diebold, Francis X., and Robert S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253–265.

Duan, Jin-Chuan, and Chung-Ying Yeh, 2010, Jump and volatility risk premiums implied by vix, *Journal of Economic Dynamics and Control* 34, 2232–2244.

Duan, Jin-Chuan, and Chung-Ying Yeh, 2011, Price and volatility dynamics implied by the vix term structure, Technical report, NUS RMI Working Paper No. 11/05.

Duffie, Darrell, Jun Pan, and Kenneth J. Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.

Durham, Garland B., 2013, Risk-neutral modeling with affine and nonaffine models, *Journal of Financial Econometrics* 11, 650–681.

Egloff, Daniel, Markus Leippold, and Liuren Wu, 2010, Valuation and optimal investing in variance swaps, *Journal of Financial and Quantitative Analysis* 45, 1279–1310.

Eraker, Bjorn, 2004, Do stock prices and volatility jump? Reconciling evidence from spot and option prices, *Journal of Finance* 59, 1367–1404.

Fang, Faming, and Cornelis W. Oosterlee, 2008, A novel pricing method for European options based on Fourier-cosine series expansions, *SIAM Journal on Scientific Computing* 31, 826–848.

Ferriani, Fabrizio, and Sergio Pastorello, 2012, Estimating and testing non-affine option pricing models with a large unbalanced panel of options, *The Econometrics Journal* 15, 171–203.

Gatheral, Jim, 2008, Consistent modeling of spx and vix options, The Fifth World Congress of the Bachelier Finance Society, London, July 2008.

Gordon, Neil J., David J. Salmond, and Adrian F. M. Smith, 1993, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *Radar and Signal Processing, IEE Proceedings F* 140, 107–113.

Hansen, Nikolaus, and Andreas Ostermeier, 1996, Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation, *Proceedings of the 1996 IEEE Conference on Evolutionary Computation (ICEC 96)* 312–317.

Jacod, Jean, and Viktor Todorov, 2010, Do price and volatility jump together?, *Annals of Applied Probability* 20, 1425–1469.

Jiang, George J., and Yisong S. Tian, 2007, Extracting model-free volatility from option prices: An examination of the VIX index, *The Journal of Derivatives* 14, 35–60.

Johannes, Michael S., Nicholas G. Polson, and Jonathan R. Stroud, 2009, Optimal filtering of jump diffusions: Extracting latent states from asset prices, *Review of Financial Studies* 22, 2759–2799.

Jones, Christopher S., 2003, The dynamics of stochastic volatility: Evidence from underlying and options markets, *Journal of Econometrics* 116, 181–224.

Kaeck, Andreas, and Carol Alexander, 2012, Volatility dynamics for the s&p 500: Further evidence from non-affine, multi-factor jump diffusions, *Journal of Banking & Finance* 36, 3110 – 3121.

Kloeden, Peter E., and Eckhard Platen, 1992, *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, Berlin, Germany).

Lindström, Erik, Jonas Ströjby, Mats Brodén, Magnus Wiktorsson, and Jan Holst, 2008, Sequential calibration of options, *Computational Statistics & Data Analysis* 52, 2877–2891.

Mencía, Javier, and Enrique Sentana, 2013, Valuation of vix derivatives, *Journal of Financial Economics* 108, 367 – 391.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Pan, Jun, 2002, The jump-risk premia implicit in options: Evidence from an integrated time-series study, *Journal of Financial Economics* 63, 3–50.

Papanicolaou, A., and R. Sircar, 2013, A regime-switching heston model for vix and s&p 500 implied volatilities, *Quantitative Finance* Forthcoming.

Pitt, Michael K, 2002, Smooth particle filters for likelihood evaluation and maximisation, Technical Report 651, The Warwick Economics Research Paper Series.

Pitt, Michael K., and Neil Shephard, 1999, Filtering via simulation: Auxiliary Particle Filters, *Journal of the American Statistical Association* 94, 590–599.

Rebonato, Riccardo, and Teresa Cardoso, 2004, Unconstrained fitting of implied volatility surfaces using a mixture of normals, *Journal of Risk* 7, 55–74.

Sepp, Artur, 2008a, Pricing options on realized variance in the heston model with jumps in returns and volatility, *Journal of Computational Finance* 11, 33–70.

Sepp, Artur, 2008b, VIX option pricing in a jump-diffusion model, *Risk Magazine* 84–89.

Song, Zhaogang, and Dacheng Xiu, 2012, A tale of two option markets: Pricing kernels and volatility risk, Technical report, Chicago Booth Research Paper No 12-10 - Fama-Miller Working Paper.

Storn, Rainer, 1996, On the usage of differential evolution for function optimization, in *Biennial Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, 519–523.

Todorov, Viktor, 2010, Variance risk premium dynamics: The role of jumps, *Review of Financial Studies* 23, 345–383.

Todorov, Viktor, and George Tauchen, 2011, Volatility jumps, *Journal of Business and Economic Statistics* 29, 356–371.

Appendix

Appendix A. Affine dependence of the VIX^2 on v_t and m_t

The expressions for the coefficients α_{VIX^2}, β_{VIX^2} and γ_{VIX^2} in Proposition II.1 are given by

$$\alpha_{\text{VIX}^2} = \left(1 + 2\lambda_1^{Yv} C\right) A, \tag{A.1}$$

$$\beta_{\text{VIX}^2} = \left(1 + 2\lambda_1^{Yv} C\right) B + \left(2\lambda_2^{Yv} C\right) \hat{A}, \tag{A.2}$$

$$\gamma_{\text{VIX}^2} = 2\lambda_0^{Yv} C + \left(1 + 2\lambda_1^{Yv} C\right) G + \left(2\lambda_2^{Yv} C\right) \hat{B}, \tag{A.3}$$

where $A = \frac{1}{a_v \tau_{\text{VIX}}} \left(e^{a_v \tau_{\text{VIX}}} - 1\right)$ and $A = 1$, if $a_v = 0$, and $C := (\theta_Z(1,0,0) - \frac{\partial \theta_Z}{\partial \phi_Y}(0,0,0) - 1)$. We can calculate the remaining coefficients:

$$B = \begin{cases} \frac{1}{\tau_{\text{VIX}}} \frac{h_v}{(a_m - a_v)} \left[\left(\frac{e^{a_m \tau_{\text{VIX}} - 1}}{a_m} \right) - \left(\frac{e^{a_v \tau_{\text{VIX}} - 1}}{a_v} \right) \right], \text{ if } a_m, a_v \neq 0, \\ \frac{h_v}{a_v} \left(e^{a_v \tau_{\text{VIX}}} - \frac{1}{a_v \tau_{\text{VIX}}} \left(e^{a_v \tau_{\text{VIX}}} - 1 \right) \right), \text{ if } a_v = a_m \neq 0, \\ \frac{h_v}{a_m} \left(\frac{1}{a_m \tau_{\text{VIX}}} \left(e^{a_m \tau_{\text{VIX}}} - 1 \right) - 1 \right), \text{ if } a_m \neq a_v = 0, \\ \frac{h_v}{a_v} \left(\frac{1}{a_v \tau_{\text{VIX}}} \left(e^{a_v \tau_{\text{VIX}}} - 1 \right) - 1 \right), \text{ if } a_v \neq a_m = 0, \\ \frac{1}{2} \tau_{\text{VIX}} h_v, \text{ if } a_m = a_v = 0, \end{cases} \tag{A.4}$$

$$G = \begin{cases} \frac{b_v}{a_v} \left[\left(\frac{e^{a_v \tau_{\text{VIX}} - 1}}{a_v \tau_{\text{VIX}}} \right) - 1 \right] - b_m B, \text{ if } a_v \neq 0, \\ \frac{b_v}{a} \left[\left(\frac{e^{a \tau_{\text{VIX}} - 1}}{a \tau_{\text{VIX}}} \right) - 1 \right] - b_m B, \text{ if } a_v = a_m \neq 0, \\ \frac{1}{2} b_v \tau_{\text{VIX}} - b_m B, \text{ if } a_m \neq a_v = 0, \\ \frac{c_m}{a_v} \left(B - \frac{1}{2} h_v \tau_{\text{VIX}} \right) + \frac{1}{a_v} \frac{\partial \theta_Z}{\partial \phi_v}(0,0,0) \lambda_0^{Yv} \left[\left(\frac{e^{a_v \tau_{\text{VIX}} - 1}}{a_v \tau_{\text{VIX}}} \right) - 1 \right], \text{ if } a_v \neq a_m = 0, \\ \frac{1}{2} \tau_{\text{VIX}} \left[\frac{\partial \theta_Z}{\partial \phi_v}(0,0,0) \lambda_0^{Yv} + c_m h_v \frac{\tau_{\text{VIX}}}{3} \right], \text{ if } a_m = a_v = 0, \end{cases} \tag{A.5}$$

$$\hat{A} = \begin{cases} \frac{e^{a_m \tau_{\text{VIX}} - 1}}{a_m \tau_{\text{VIX}}}, \text{ if } a_m \neq 0, \\ 1, \text{ if } a_m = 0, \end{cases} \qquad \hat{B} = \begin{cases} b_m \left(1 - \hat{A} \right), \text{ if } a_m \neq 0, \\ \frac{c_m \tau_{\text{VIX}}}{2}, \text{ if } a_m = 0. \end{cases} \tag{A.6}$$

Furthermore,

$$\begin{aligned} a_m &= \left(\frac{\partial \theta_Z}{\partial \phi_m}(0,0,0) \lambda_1^m - \kappa_m \right), \quad c_m = \left(\kappa_m \theta_m + \frac{\partial \theta_Z}{\partial \phi_m}(0,0,0) \lambda_0^m \right), \\ a_v &= \left(\frac{\partial \theta_Z}{\partial \phi_v}(0,0,0) \lambda_1^{Yv} - \kappa_v \right), \quad h_v = \left(\kappa_v + \frac{\partial \theta_Z}{\partial \phi_v}(0,0,0) \lambda_2^{Yv} \right), \end{aligned} \tag{A.7}$$

and if $a_m \neq 0$:

$$b_m = -\frac{c_m}{a_m}, \quad b_v = b_m \left(\kappa_v + \frac{\partial \theta_Z}{\partial \phi_v}(0,0,0)\lambda_2^{Yv} \right) + \frac{\partial \theta_Z}{\partial \phi_v}(0,0,0)\lambda_0^{Yv}. \tag{A.8}$$

Appendix B. Characteristic functions

Assuming that the characteristic functions of the processes Y, VIX^2 take the exponential affine form

$$\Psi_{VIX_T^2}(t,v,m;\omega) = \mathbb{E}_t\left[e^{\omega VIX_T^2}\right] = e^{\alpha(T-t)+\beta(T-t)v+\gamma(T-t)m},$$
$$\Psi_{Y_T}(t,v,m;\omega) = \mathbb{E}_t\left[e^{\omega Y_T}\right] = e^{\alpha_Y(T-t)+\beta_Y(T-t)y+\gamma_Y(T-t)v+\delta_Y(T-t)m},$$

with $\omega \in \mathbb{C}$, the coefficients in the definition of $\Psi_{VIX_T^2}$ satisfy the following ODEs:[23]

$$\begin{aligned}
&-\alpha'(T-t) + \gamma(T-t)\kappa_m\theta_m + \lambda_0^{Yv}\left(\theta_Z(0,\beta(T-t),0)-1\right) + \lambda_0^m\left(\theta_Z(0,0,\gamma(T-t))-1\right) = 0\\
&-\beta'(T-t) - \beta(T-t)\kappa_v + \frac{1}{2}\sigma_v^2\beta^2(T-t) + \lambda_1^{Yv}\left(\theta_Z(0,\beta(T-t),0)-1\right) = 0\\
&-\gamma'(T-t) - \gamma(T-t)\kappa_m + \frac{1}{2}\sigma_m^2\gamma^2(T-t) + \kappa_v\beta(T-t) + \lambda_2^{Yv}\left(\theta_Z(0,\beta(T-t),0)-1\right) +\\
&\qquad \lambda_1^m\left(\theta_Z(0,0,\gamma(T-t))-1\right) = 0
\end{aligned}$$

$\forall t \in (0,T]$, with boundary conditions $\alpha(0)=0$, $\beta(0)=\omega_1$ and $\gamma(0)=\omega_2$, where $\omega_1 := \omega\alpha_{VIX^2}$ and $\omega_2 := \omega\beta_{VIX^2}$ (the coefficients α_{VIX^2} and β_{VIX^2} are defined in Appendix A). The coefficients of

[23] This relies on the fact that the Poisson processes driving the jumps in v_t and in m_t are independent.

Ψ_{Y_T} satisfy the following ODEs for $t \in (0, T]$:

$$
\begin{aligned}
&- \alpha_Y'(T-t) + \beta_Y(T-t)(-\lambda_0^{Yv}(\theta_Z(1,0,0)-1)) + \delta_Y(T-t)\kappa_m\theta_m \\
&\qquad + \lambda_0^{Yv}[\theta_Z(\beta_Y(T-t), \gamma_Y(T-t), 0) - 1] + \lambda_0^m[\theta_Z(0,0,\delta_Y(T-t)) - 1] = 0 \\
&- \beta_Y'(T-t) = 0 \\
&- \gamma_Y'(T-t) - \beta_Y(T-t)\lambda_1^{Yv}(\theta_Z(1,0,0)-1) - \frac{1}{2}\beta_Y(T-t) - \gamma_Y(T-t)\kappa_v + \frac{1}{2}\beta_Y(T-t)^2 \\
&\qquad + \frac{1}{2}\gamma_Y(T-t)^2\sigma_v^2 + \beta_Y(T-t)\gamma_Y(T-t)\sigma_v\rho_{Y,v} + \lambda_1^{Yv}[\theta_Z(\beta_Y(T-t), \gamma_Y(T-t), 0) - 1] = 0 \\
&- \delta_Y'(T-t) - \beta_Y(T-t)\lambda_2^{Yv}(\theta_Z(1,0,0)-1) + \gamma_Y(T-t)\kappa_v - \delta_Y(T-t)\kappa_m + \frac{1}{2}\delta_Y(T-t)^2\sigma_m^2 \\
&\qquad + \lambda_2^{Yv}[\theta_Z(\beta_Y(T-t), \gamma_Y(T-t), 0) - 1] + \lambda_1^m[\theta_Z(0,0,\delta_Y(T-t)) - 1] = 0,
\end{aligned}
$$

with boundary conditions $\alpha_Y(0) = 0$, $\beta_Y(0) = \omega$, $\gamma_Y(0) = 0$ and $\delta_Y(0) = 0$.

Appendix C. Coefficients for the Fourier cosine expansion

Here we give the expression for $U_n^{\text{VIX}^2}$, the Fourier cosine transform of the VIX options' payoff. To ease notation, we drop the subscript VIX for $a_{\text{VIX}}, b_{\text{VIX}}$ and define $\omega_n := \frac{n\pi}{b-a}$. For $n > 0$, we obtain

$$
\begin{aligned}
U_n^{\text{VIX}^2} &= \int_a^b \left(\sqrt{x} - K\right)^+ \cos\left(\omega_n(x-a)\right) dx \qquad \text{(C.1)} \\
&= \frac{2}{b-a} Re\left\{ e^{-i\omega_n a}\left[\frac{\sqrt{b}e^{-i\omega_n b}}{i\omega_n} + \frac{\sqrt{\pi}}{2(-i\omega_n)^{3/2}}\left(\text{erfz}(\sqrt{-i\omega_n b}) - \text{erfz}(K\sqrt{-i\omega_n})\right)\right]\right\}.
\end{aligned}
$$

where erfz is the error function for complex numbers.

For $n = 0$,

$$
U_0^{\text{VIX}^2} = \frac{2}{b-a}\left[\frac{2}{3}b^{3/2} - Kb + \frac{1}{3}K^3\right]. \qquad \text{(C.2)}
$$

Appendix D. Particle filter

The log-likelihood of a time series of $n+1$ observations with joint density p, conditional on a set of parameters Θ, is equal to

$$\log p(y^{t_n}|\Theta) = \log p(y_{t_0}, ..., y_{t_n}|\Theta) = \sum_{k=1}^{n} \log p(y_{t_k}|y^{t_{k-1}}, \Theta) + \log p(y_{t_0}|\Theta) \tag{D.1}$$

where, by the Law of Total Probability,

$$p(y_{t_k}|y^{t_{k-1}}, \Theta) = \int p(y_{t_k}|L_{t_k}, \Theta) p(L_{t_k}|y^{t_{k-1}}, \Theta) dL_{t_k}. \tag{D.2}$$

Given an initial density $p(L_{t_0}|\Theta)$, the transition density of the state variables $p(L_{t_k}|L_{t_{k-1}}, \Theta)$ and the likelihood function $p(y_{t_k}|L_{t_k}, \Theta)$, filtering methods allow us to estimate the distribution $p(L_{t_k}|y^{t_k}, \Theta)$ of the current state at time $t_k = k\Delta t$, given all observations up to that time. In the following, we simplify notation and drop the subscript for the conditioning on the parameters Θ. The filtering density is given by Bayes's formula,

$$p(L_{t_k}|y^{t_k}) \propto p(y_{t_k}|L_{t_k}) p(L_{t_k}|y^{t_{k-1}}). \tag{D.3}$$

The likelihood function is known, but the predictive distribution of the state is not. It is given by the following integral, which involves the previous filtering density.

$$p(L_{t_k}|y^{t_{k-1}}) = \int p(L_{t_k}|L_{t_{k-1}}) p(L_{t_{k-1}}|y^{t_{k-1}}) dL_{t_{k-1}} \tag{D.4}$$

The key idea is to approximate the posterior density function of the latent variables $p(L_{t_k}|y^{t_k})$ by a sum of point masses positioned at strategic points, called particles, $\{L_{t_k}^{(i)}\}_{1 \le i \le n_p}$:

$$\hat{p}(L_{t_k}|y^{t_k}) = \sum_{i=1}^{n_p} \pi_{t_k}^{(i)} \delta(L_{t_k} - L_{t_k}^{(i)}), \tag{D.5}$$

where $\pi_{t_k}^{(i)}$ denotes the normalized importance weight for particle i, $\delta(.)$ is the Dirac delta function, and n_p is the number of support points (particles) for $\hat{p}(L_{t_k}|y^{t_k})$. Then, we can recursively calculate

the filtering density by

$$\begin{aligned}\hat{p}(L_{t_k}|y^{t_k}) &\propto \int p(y_{t_k}|L_{t_k})p(L_{t_k}|L_{t_{k-1}})\hat{p}(L_{t_{k-1}}|y^{t_{k-1}})dL_{t_{k-1}} \\ &= \sum_{i=1}^{n_p} p(y_{t_k}|L_{t_k})p(L_{t_k}|L_{t_{k-1}}^{(i)})\pi_{t_{k-1}}^{(i)}. \end{aligned} \tag{D.6}$$

To implement the particle filter, ones needs to be able to simulate at every time t_k a number n_p of particles $L_{t_k}^{(i)}, i = 1, ..., n_p$ from $p(L_{t_k}|y^{t_{k-1}})$ and to be able to evaluate $p(y_{t_k}|L_{t_k}^{(i)})$. Based on these simulated particles, we can approximate $p(y_{t_k}|y^{t_{k-1}})$ by

$$p(y_{t_k}|y^{t_{k-1}}) \approx \frac{1}{n_p}\sum_{i=1}^{n_p} p(y_{t_k}|L_{t_k}^{(i)}). \tag{D.7}$$

The available observations are the S&P 500 daily returns, the VIX levels, and the option prices on both indexes. In the following, the notation is as in Section IV.B. We used $n_p = 15,000$ particles on days when the observations contain option prices and $n_p = 8,000$ when the observations are only composed of the S&P500 returns and VIX index levels. Larger numbers of particles did not change our estimates, but increased the computational burden.

The filtering algorithm can be decomposed into the following steps.

Step 1: Initialization. We simulate n_p initial particles for the latent variables $\{v_{t_0}^{(i)}, m_{t_0}^{(i)}\}_{i=1,...,n_p}$ which are compatible with the initial value of the VIX squared, i.e., given the specification (23). The following steps are repeated for each time step t_k in the grid from $k = 0$ to $k = M - 1$.

Step 2: First-stage resampling. At this point, we assume that we have n_p particles (i.e., possible values of m_t and v_t) at time t_k given all observations y^{t_k} up to t_k. At time t_{k+1}, there are new observations y_{k+1}. The goal of this step is to retain, from the previous sample of particles $\{v_{t_k}^{(i)}, m_{t_k}^{(i)}\}_{1\leq i\leq n_p}$, only those which are likely to generate the new observations y_{k+1}. For this purpose, we assign a weight (the so called "first-stage weights") to each particle, which is proportional to the likelihood of new market observations $y_{t_{k+1}}$ given the value of the particle L_{t_k} at time t_k. Intuitively, particles that are compatible with the new observations will be assigned larger weights

than other particles. To increase the speed of the first-stage resampling, we do not consider options as part of the observations $y_{t_{k+1}}$ (only in this step) and limit $y_{t_{k+1}}$ to the values of the indexes.

The first stage weight $\omega_{t_{k+1}}^{(i)}$ assigned to the i^{th} particle $L_{t_k}^{(i)}$ at time t_{k+1} is given by

$$\omega_{t_{k+1}}^{(i)} = p(L_{t_k}^{(i)}|y_{t_{k+1}}) \propto p(y_{t_{k+1}}|L_{t_k}^{(i)})$$

where $p(y_{t_{k+1}}|L_{t_k}^{(i)})$ is the density of the observation vector $y_{t_{k+1}}$ given the values of the particle vector $L_{t_k}^{(i)}$. The importance weights $\{\omega_{t_{k+1}}^{(i)}\}_{1\leq i\leq n_p}$ add up to 1, so that they define a proper probability mass function. Conditioning on the number of jumps in ΔY_t (or equivalently in Δv_t) and Δm_t gives

$$\omega_{t_{k+1}}^{(i)} \propto \sum_{j,l\in\mathbb{N}} p(y_{t_{k+1}}|L_{t_k}^{(i)}, \Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m})\mathbb{P}(\Delta N_{t_k}^{Yv} = j, \Delta N_{t_k}^{m} = l).$$

Given that we use daily observations, we limit the possible number of jumps of the Poisson random variables $\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m}$ to zero or one (this Bernoulli approximation is found to be very accurate in Johannes et al. (2009)). We recall that the new observation is composed of the S&P 500 returns and the VIX level $y_{t_{k+1}} = (\Delta Y_{t_k} = Y_{t_{k+1}} - Y_{t_k}, \text{VIX}_{t_{k+1}}^2)$. Since the $\text{VIX}_{t_{k+1}}^2$ is a sum of normal distributions and no more than two exponential distributions, there is no closed form for this bivariate density in the general case. To preserve tractability, we approximate the exponentially distributed jump sizes by a categorical distribution (a generalization of a Bernoulli distribution) which is supported in a certain number of (the corresponding exponential distribution's) quantiles.[24] As a consequence, the weight $\omega_{t_{k+1}}^{(i)}$ is a sum of weighted bivariate normal densities.

To eliminate the particles $\{L_{t_k}^{(i)}\}_{1\leq i\leq n_p}$ that are not likely to generate the new observations $y_{t_{k+1}}$, we resample (with replacement) the particles according to a stratified resampling scheme:[25]

$$z(i) \sim StratRes(n_p, \omega_{t_{k+1}}^{(1)}, ..., \omega_{t_{k+1}}^{(n_p)}).$$

This makes it possible to create a new sample of n_p latent factors $\{L_{t_k}^{(j)}\}_{1\leq j\leq n_p}$ which are now equally likely. Indeed, particles representing m_{t_k} and v_{t_k} are shuffled into a new set of particles:

[24] Robustness tests were performed on simulated data to check that the choice of quantiles was appropriate.

[25] We checked that using a multinomial or stratified resampling scheme gives similar results.

$\{m_{t_k}^{(j)}, v_{t_k}^{(j)}\}_{j=1..n_p} = \{m_{t_k}^{z(i)}, v_{t_k}^{z(i)}\}_{i=1..n_p}$. We resample the same number of particles, although this is in principle not necessary.

The next step of the particle filter consists in propagating the latent factors according to their conditional density given the previous values $L_{t_k}^{(i)}$ and the new observations $y_{t_{k+1}}$:

$$L_{t_{k+1}}^{(i)} \sim p(L_{t_{k+1}}|L_{t_k}^{(i)}, y_{t_{k+1}}).$$

Because the distribution $p(L_{t_{k+1}}|L_{t_k}^{(i)}, y_{t_{k+1}})$ is not known in closed form, we use a proposal density $q(L_{t_{k+1}}|L_{t_k}^{(i)}, y_{t_{k+1}})$. Propagating v_t and m_t requires preliminary knowledge on the jump components, so we first focus on the jumps.

Step 3: Generating the jumps. We calculate the joint probability of jumps in ΔY_t (or equivalently in Δv_t) and Δm_t between t_k and t_{k+1} using

$$\mathbb{P}(\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m}|y_{t_{k+1}}) \propto p(y_{t_{k+1}}|\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m})\mathbb{P}(\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m}). \tag{D.8}$$

Conditionally on the jump sizes in v_t and m_t, the first part of the right-hand side has already been calculated in the first-stage weights. Using Bayes's rule, we get an approximation for $\mathbb{P}(\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{m}|y_{t_{k+1}})$.

We infer the jump size in the returns following Johannes et al. (2009): $Z_{t_k}^{Y(i)}|\Delta N_{t_k}^{Yv}, y_{t_{k+1}}$ is normally distributed $\mathcal{N}(\mu_J^{Y(i)}, \sigma_J^{Y(i)})$ where $\mu_J^{Y(i)}$ and $\sigma_J^{Y(i)}$ are given by

$$\left(\sigma_J^{Y(i)}\right)^2 = \left(\frac{1}{\Delta N_{t_k}^{Yv(i)}(\sigma_Y^{\mathbb{P}})^2} + \frac{1}{\hat{v}_{t_{k+1}}^{(i)}}\right)^{-1}$$
$$\mu_J^{Y(i)} = \left(\sigma_J^{Y(i)}\right)^2 \frac{Y_{t_{k+1}} - \tilde{\mu}_Y^{(i)}}{\hat{v}_{t_{k+1}}^{(i)}} + \frac{(\sigma_J^{Y(i)})^2}{(\sigma_Y^{\mathbb{P}})^2}\mu_Y^{\mathbb{P}}$$

where $\hat{v}_{t_{k+1}}^{(i)}$ is an estimate of $v_{t_{k+1}}$ given the information we have up to time t_k and particle i; we use $\hat{v}_{t_{k+1}}^{(i)} = \mathbb{E}[v_{t_{k+1}}|v_{t_k}^{(i)}]$ and

$$\tilde{\mu}_Y^{(i)} = -\left(\lambda^{Yv}(\theta_Z(1,0,0) - 1) + \frac{1}{2}v_{t_k}^{(i)} - \Delta N_{t_k}^{Yv(i)}\right)\Delta t.$$

Finally, we simulate the jump sizes for Δv_t and Δm_t according to their exponential law.

Step 4: Propagating the volatility and central tendency. The latent factors v and m are propagated following a Milstein discretization scheme of the SDE. We use the full truncation method to prevent them from taking negative values.

Step 5: Computing the filtering density. At this point, the newly generated particles $\{L^{(i)}_{t_{k+1}}\}_{1\leq i\leq n_p}$ are a sample of $p(L_{t_{k+1}}|y^{t_{k+1}})$. We now calculate the *second-stage weights* $\{\pi^{(i)}_{t_{k+1}}\}_{1\leq i\leq n_p}$ which approximate the probabilities $p(L^{(i)}_{t_{k+1}}|y^{t_{k+1}})$ and give an approximation for the filtering density at time t_{k+1}. These weights are proportional to the likelihood of observations at time t_{k+1} given the propagated particles $L^{(i)}_{t_{k+1}}$, with a correction related to the proposal density

$$\pi^{(i)}_{t_{k+1}} \propto \frac{p(L^{(i)}_{t_{k+1}}|L^{(i)}_{t_k})p(y_{t_{k+1}}|L^{(i)}_{t_{k+1}})}{\omega^{z(i)}_{t_{k+1}} q(L^{(i)}_{t_{k+1}}|L^{(i)}_{t_k}, y_{t_{k+1}})}.$$

The posterior distribution of the state variables is approximated by

$$\hat{p}(L_{t_{k+1}}|y^{t_{k+1}}) = \sum_{i=1}^{n_p} \pi^{(i)}_{t_{k+1}} \delta(L_{t_{k+1}} - L^{(i)}_{t_{k+1}}).$$

We choose the most likely value of a given factor by taking the expectation of the estimated filtering density, e.g., $\hat{v}_{t_{k+1}} = \mathbb{E}_{\hat{p}}[v^{(i)}_{t_{k+1}}]$.

The algorithm described above extracts latent factors if one assumes that the model parameters are known. Pitt (2002) builds on Gordon et al. (1993) to show that the parameters can be estimated using the Maximum Likelihood Importance Sampling Criterion, defined as the product over time of the averages of the second-stage weights. The likelihood of observations given the values of the particles is then estimated by the average of the second-stage weights over the particles

$$\hat{p}(y^{t_n}|\Theta, \mathcal{M}) = \prod_{k=1}^{M} \hat{p}(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M})\hat{p}(y_{t_0}|\Theta, \mathcal{M}),$$

where $\hat{p}(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M}) = \frac{1}{n_p}\sum_{i=1}^{n_p} \pi^{(i)}_{t_k}$.

Appendix E. Specific data treatment for the particle filter

The S&P 500 options dataset contains a large amount of ATM options compared to OTM and deep OTM options. If we use the filter (within the maximum likelihood procedure) on this entire dataset, the fitting of ATM options will be its priority rather than (deep) OTM options. Given the formula of Breeden and Litzenberger (1978), this results in fitting the body of the S&P 500 returns distribution rather than the tails, which is not what we want. We need information about the extreme events contained in the data to be incorporated into the models. For this reason, we interpolate the S&P 500 IV slices and re-sample the option prices from the resulting parametric fit uniformly with respect to moneynesses.[26] Other advantages of our use of interpolation is that the resulting data is arbitrage free, we have fewer points for each slice (but still accurately representing the information of each slice), thus reducing the computational complexity.[27]

For the interpolation, we use the efficient mixture of log-normals approach of Rebonato and Cardoso (2004) to have a parametric fit for each S&P 500 implied volatility slice. The RMSE of the S&P 500 implied volatilities parametric fits are on average around 0.25% and we therefore do not lose information, especially given the market bid–ask spread. Finally, using the parametric fit, we can sample a fixed number (we have chosen 15) of "market option prices" for the desired strikes. We have chosen to resample the option prices from each parametric slice uniformly in the strike (or, equivalently, the moneyness). We however do not resample the options for which the strike is smaller than 40% or larger than 140% of the current futures price. The reason is that there are usually only one or two options outside this interval of moneyness and we do not wish to re-sample options where the interpolation results could be driven by an outlier.

We do not perform any interpolation for the VIX options dataset, as most VIX options are OTM and therefore contain information about the tails of the VIX distribution (i.e., variance and central tendency processes).

As the datasets comprise a large number of options (up to 600 a day), it is unfeasible to calculate

[26] It is common to interpolate data, see, e.g., Broadie et al. (2007). This eliminates arbitrage opportunities in the data and removes the accumulation of options around the ATM region.

[27] Since we have considered mid-prices and because of synchronization issues between the underlying and the options, implied volatility slices are not guaranteed to be arbitrage free.

the option prices every day for every particle. As a consequence, we follow Pan (2002) and Johannes et al. (2009), among others, and use weekly (Wednesday) options data. Furthermore, this eliminates beginning-of-week and end-of-week effects. Our particle filter uses daily time steps and incorporates information on the underlying indexes on a daily basis (i.e., only options are considered weekly).

We decompose the time series of observations into two periods. The first period is from March 1st, 2006 to Feb 28, 2009 (shortly after the VIX index increased to its highest point). This was a rather calm period,[28] that we will use as the in-sample estimation period. Our out-of-sample period starts on March 1st, 2009 and ends on October 29, 2010. This period includes very high levels of volatility (for implied volatilities from the S&P 500 and VIX options as well as the VIX index values). The last column of Table IV presents the number of options within each moneyness and maturity range in both periods. In particular, our dataset contains, in the in-sample period, 608 close-to-maturity OTM call options on the S&P 500 with a moneyness larger than 1.05, and 2,243 OTM put options with a moneyness smaller than 0.95. These options have maturities shorter than two months. Analogously, in the out-of-sample period, the dataset contains 737 close-to-maturity OTM calls and 2,032 OTM put options. As highlighted in Bollerslev and Todorov (2011), these options provide valuable information on jumps as they have little value unless a large movement in the S&P 500 is possible. Similarly, OTM calls on the VIX with short maturities contain information on the extreme upwards moves in the VIX index, and help identify the heavy-tailedness of the right tail of the VIX distribution. Our dataset contains 1,006 such options on the VIX with a moneyness larger than 1.1 in the in-sample period and 1,269 in the out-of-sample period.

[28] We have decided to include the beginning of the financial crisis so that the in-sample period actually includes several dates with extreme events.

Table I: Descriptive statistics for daily S&P 500 futures log-returns and daily VIX levels for the periods from March 2006 to February 2009, and March 2009 to October 2010. We report the mean (Mean), standard deviation (Std), skewness (Skew), and kurtosis (Kurt).

	March 2006–February 2009				March 2009–October 2010			
	Mean	Std	Skew	Kurt	Mean	Std	Skew	Kurt
S&P 500	-0.0007	0.0159	0.0722	14.0772	0.0007	0.0158	-0.3283	6.8683
VIX	0.2044	0.1208	2.6560	10.9620	0.2907	0.1025	1.1929	3.8631

Table II: $\mathbb{P}$-parameters and measure-independent parameters, estimated using the particle filter for the different models and datasets. The standard errors are in italics below each parameter, except for θ_m^P, which is calculated as $\theta_m^P = \theta_m^Q \kappa_m^Q / \kappa_m^P$. $\mathcal{D}_1$ contains S&P 500 and VIX levels. $\mathcal{D}_2$ (resp. $\mathcal{D}_3$) contains in addition the S&P 500 options (resp. VIX options). $\mathcal{D}_4$ contains all available data. The estimation period is from March 2006 to February 2009.

	$\mathcal{D}_1$			$\mathcal{D}_2$			$\mathcal{D}_3$			$\mathcal{D}_4$		
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2
	$\mathbb{P}$ and $\mathbb{Q}$ parameters											
λ_0^{Yv}	0.400	0.050	–	0.010	0.050	–	0.160	0.060	–	0.170	0.088	–
	0.204	*0.082*		*0.056*	*0.026*		*0.054*	*0.042*		*0.099*	*0.022*	
λ_1^{Yv}	1.000	3.350	–	1.100	2.820	–	0.240	2.850	–	0.150	2.376	–
	0.529	*1.896*		*0.656*	*0.183*		*0.161*	*0.296*		*0.485*	*0.504*	
λ_2^{Yv}	3.000	–	–	4.250	–	–	3.640	–	–	4.000	–	–
	0.637			*0.161*			*0.854*			*1.208*		
σ_m	0.080	–	0.110	0.240	–	0.190	0.350	–	0.300	0.190	–	0.280
	0.173		*0.135*	*0.088*		*0.054*	*0.117*		*0.080*	*1.034*		*0.104*
σ_v	0.840	0.550	0.840	0.720	0.510	0.910	0.770	0.390	0.770	0.790	0.375	0.830
	0.129	*0.195*	*0.142*	*0.064*	*0.072*	*0.058*	*0.069*	*0.061*	*0.041*	*0.192*	*0.034*	*0.119*
ρ_{Yv}	-0.620	-0.900	-0.620	-0.800	-0.890	-0.790	-0.780	-0.890	-0.760	-0.860	-0.934	-0.820
	0.138	*0.227*	*0.152*	*0.105*	*0.023*	*0.023*	*0.074*	*0.050*	*0.184*	*0.028*	*0.051*	*0.086*
m_0		0.030			0.026			0.024			0.023	
		0.008			*0.002*			–			*0.003*	
	$\mathbb{P}$ parameters											
$\kappa_v^{\mathbb{P}}$	7.150	11.400	9.300	6.850	7.920	7.240	8.200	7.500	7.620	6.800	7.558	7.400
	2.540	*3.476*	*1.576*	*0.881*	*0.384*	*0.697*	*0.696*	*0.220*	*0.341*	*0.628*	*0.531*	*0.465*
$\kappa_m^{\mathbb{P}}$	0.050	–	0.080	0.280	–	0.200	0.300	–	0.340	0.300	–	0.300
	0.489		*0.464*	*0.258*		*0.224*	*0.178*		*0.092*	*0.289*		*0.596*
$\theta_m^{\mathbb{P}}$	0.011	–	0.109	0.045	–	0.063	0.109	–	0.109	0.033	–	0.016
$\nu_v^{\mathbb{P}}$	0.030	0.090	–	0.020	0.160	–	0.220	0.150	–	0.030	0.204	–
	0.155	*0.114*		*0.152*	*0.034*		*0.054*	*0.043*		*0.113*	*0.105*	
η_Y	0.850	0.100	0.900	0.700	0.500	0.500	0.340	0.450	0.430	0.500	0.782	0.800
	0.190	*0.458*	*0.234*	*0.158*	*0.112*	*0.129*	*0.282*	*0.221*	*0.235*	*0.255*	*0.181*	*0.163*

Table III: $\mathbb{Q}$-parameters estimated by the particle filter for the different models and datasets. $\mathcal{D}_1$ contains S&P 500 and VIX levels. $\mathcal{D}_2$ (resp. $\mathcal{D}_3$) contains also the S&P 500 options (resp. VIX options). $\mathcal{D}_4$ contains all available data. The estimation period is from March 2006 to February 2009. The resulting log-likelihood is shown, as well as the Akaike and Bayes Information Criteria.

	$\mathcal{D}_1$			$\mathcal{D}_2$			$\mathcal{D}_3$			$\mathcal{D}_4$		
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2
						$\mathbb{Q}$ **parameters**						
κ_v	12.000	8.300	12.000	5.450	4.630	6.860	6.320	4.080	7.160	5.700	4.423	5.900
	3.807	*1.690*	*3.109*	*0.374*	*0.341*	*0.617*	*0.427*	*0.396*	*0.663*	*1.034*	*0.415*	*1.034*
κ_m	0.490	–	0.500	0.210	–	0.400	0.230	–	0.300	0.200	–	0.480
	0.271		*0.251*	*0.076*		*0.104*	*0.121*		*0.110*	*0.192*		*0.111*
θ_m	0.110	–	0.040	0.060	–	0.020	0.080	–	0.150	0.050	–	0.010
	0.076		*0.136*	*0.018*		*0.012*	*0.036*		*0.074*	*0.028*		*0.030*
ν_v	0.050	0.130	–	0.030	0.180	–	0.180	0.220	–	0.030	0.128	–
	0.236	*0.374*		*0.081*	*0.038*		*0.075*	*0.043*		*0.001*	0.058	
μ_Y	-0.010	-0.050	–	-0.110	-0.110	–	-0.060	-0.090	–	-0.060	-0.118	–
	0.060	*0.124*		*0.062*	*0.038*		*0.026*	*0.024*		*0.023*	*0.037*	
σ_Y	0.120	0.033	–	0.110	0.160	–	0.140	0.100	–	0.100	0.114	–
	0.108	*0.291*		*0.032*	*0.026*		*0.022*	*0.023*		*0.021*	*0.023*	
η_v	4.850	-3.100	2.700	-1.400	-3.290	-0.380	-1.880	-3.420	-0.460	-1.100	-3.135	-1.500
η_m	0.440	–	0.420	-0.070	–	0.200	-0.070	–	-0.040	-0.100	–	0.180
Log-likelihood	11376	11319	11274	6764	1439	6279	9404	7791	8060	6827	1484	4120
AIC	-22714	-22610	-22528	-13486	-2844	-12532	-18762	-15544	-16090	-13600	-2924	-8204
BIC	-22713	-22547	-22483	-13392	-2768	-12474	-18659	-15459	-16023	-13479	-2825	-8123

Table IV: Root Mean Square Relative Errors (RMSRE) for the SVJ2, SVJ and SV2 models for different datasets. Option pricing errors are reported conditional on moneyness (Mon.) and time to maturity (TTM) in months (M). The last column reports the number of options (#) in each category. Moneyness is defined as the ratio between the strike and the corresponding maturity future's price.

	$\mathcal{D}_1$			$\mathcal{D}_2$			$\mathcal{D}_3$			$\mathcal{D}_4$			
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	#
						In-sample S&P 500 options RMSREs							
Overall	1.571	0.850	1.076	0.270	0.555	0.271	0.817	0.691	0.866	0.259	0.374	0.291	14351
$Mon. < 0.7$	0.601	0.625	0.606	0.402	0.970	0.403	0.543	0.665	0.422	0.375	0.670	0.389	1365
$0.7 \leq Mon. \leq 0.95$	0.345	0.398	0.370	0.235	0.475	0.257	0.311	0.304	0.313	0.199	0.304	0.260	6498
$0.95 \leq Mon. \leq 1.05$	0.521	0.367	0.359	0.213	0.195	0.193	0.260	0.331	0.266	0.239	0.228	0.226	3673
$1.05 \leq Mon. \leq 1.2$	3.123	1.779	1.942	0.288	0.528	0.285	1.208	1.331	1.277	0.286	0.390	0.331	2287
$Mon. > 1.2$	4.528	1.396	3.409	0.450	1.294	0.387	3.063	1.509	3.340	0.468	0.697	0.494	528
$TTM \leq 2$ M	1.342	0.979	0.905	0.326	0.508	0.344	0.730	0.987	0.760	0.313	0.371	0.369	4476
2 M $\leq TTM < 6$ M	1.424	0.797	1.061	0.243	0.475	0.222	0.666	0.493	0.683	0.227	0.298	0.232	5499
$TTM > 6$ M	1.923	0.767	1.245	0.235	0.679	0.241	1.042	0.516	1.132	0.234	0.455	0.266	4376
						Out-of-sample S&P 500 options RMSREs							
Overall	1.348	0.751	1.144	0.419	0.683	0.424	1.440	0.816	1.529	0.406	0.512	0.420	13323
$Mon. < 0.7$	0.828	0.898	0.833	0.692	0.585	0.663	0.521	0.694	0.588	0.646	0.713	0.649	3046
$0.7 \leq Mon. \leq 0.95$	0.430	0.590	0.474	0.334	0.283	0.373	0.241	0.365	0.374	0.319	0.391	0.373	4755
$0.95 \leq Mon. \leq 1.05$	0.258	0.379	0.273	0.138	0.490	0.171	0.247	0.460	0.244	0.165	0.306	0.169	1962
$1.05 \leq Mon. \leq 1.2$	1.833	1.006	1.530	0.237	1.161	0.256	1.555	1.348	1.552	0.257	0.500	0.256	2349
$Mon. > 1.2$	3.305	0.772	2.668	0.385	0.976	0.394	4.136	1.313	4.421	0.434	0.609	0.395	1211
$TTM \leq 2$ M	1.122	0.992	0.954	0.440	1.089	0.504	0.991	1.255	0.992	0.461	0.553	0.514	3580
2 M $\leq TTM < 6$ M	1.330	0.609	1.182	0.403	0.441	0.384	1.478	0.630	1.557	0.373	0.452	0.378	5678
$TTM > 6$ M	1.542	0.674	1.240	0.421	0.460	0.401	1.697	0.492	1.847	0.398	0.552	0.381	4065
						In-sample VIX options RMSREs							
Overall	0.660	0.738	0.675	0.458	0.567	0.579	0.263	0.387	0.292	0.406	0.455	0.463	4148
$Mon. > 1.3$	0.836	0.764	0.800	0.591	0.610	0.696	0.408	0.429	0.464	0.548	0.514	0.585	1100
$1.1 \leq Mon. \leq 1.3$	0.660	0.811	0.743	0.474	0.624	0.654	0.205	0.414	0.222	0.397	0.482	0.505	963
$0.9 \leq Mon. \leq 1.1$	0.564	0.751	0.663	0.367	0.596	0.505	0.158	0.362	0.188	0.309	0.423	0.382	1075
$Mon. < 0.9$	0.524	0.609	0.428	0.350	0.407	0.411	0.188	0.334	0.177	0.313	0.387	0.331	1010
$TTM \leq 2$ M	0.613	0.761	0.751	0.477	0.592	0.619	0.293	0.363	0.328	0.400	0.459	0.493	1948
$TTM > 2$ M	0.699	0.717	0.601	0.440	0.544	0.541	0.233	0.407	0.256	0.411	0.450	0.435	2200
						Out-of-sample VIX options RMSREs							
Overall	0.851	0.894	0.817	0.725	0.722	0.693	0.542	0.680	0.568	0.691	0.756	0.660	5118
$Mon. > 1.3$	0.977	0.916	0.958	0.893	0.792	0.870	0.768	0.724	0.820	0.884	0.813	0.858	2039
$1.1 \leq Mon. \leq 1.3$	0.853	0.915	0.803	0.696	0.756	0.665	0.389	0.717	0.377	0.647	0.771	0.607	916
$0.9 \leq Mon. \leq 1.1$	0.733	0.873	0.700	0.569	0.708	0.528	0.279	0.652	0.273	0.499	0.716	0.459	1109
$Mon. < 0.9$	0.687	0.855	0.623	0.494	0.543	0.432	0.268	0.583	0.258	0.428	0.661	0.382	1054
$TTM \leq 2$ M	0.815	0.846	0.799	0.706	0.667	0.684	0.551	0.594	0.597	0.674	0.698	0.653	1054
$TTM > 2$ M	0.878	0.930	0.830	0.739	0.762	0.699	0.535	0.740	0.545	0.704	0.797	0.665	2237

Table V: Diebold–Mariano test values for in- and out-of-sample errors on S&P 500 and VIX option prices, for the different models and estimation datasets. Two loss functions are considered: the average Mean Square Error (MSE) and the average Mean Square Relative Error (MSRE) of the SVJ/SV2 model with respect to the SVJ2 model. Standard errors are calculated using the Newey and West (1987) estimator with the optimal numbers of lags following Andrews (1991).

	$\mathcal{D}_2$				$\mathcal{D}_3$				$\mathcal{D}_4$			
	SVJ		SV2		SVJ		SV2		SVJ		SV2	
	MSE	MSRE	MSE	MSRE	MSE	MSRE	MSE	MSRE	MSE	MSRE	MSE	MSRE
Panel A: Errors on S&P 500 option prices												
In-sample	65.437	0.134	3.728	0.172	-4.563	0.328	32.589	1.037	57.423	0.120	86.592	0.200
Out-of-sample	59.570	0.310	13.333	0.246	-32.023	-1.246	48.785	1.649	49.302	0.315	-5.166	0.134
Panel B: Errors on VIX option prices												
In-sample	0.02	0.278	<1e-3	0.410	<1e-3	0.149	<1e-3	0.360	<1e-3	0.354	<1e-3	0.382
Out-of-sample	0.02	0.484	<1e-3	-0.503	<1e-3	0.190	<1e-3	0.535	0.002	0.451	<1e-3	-0.409

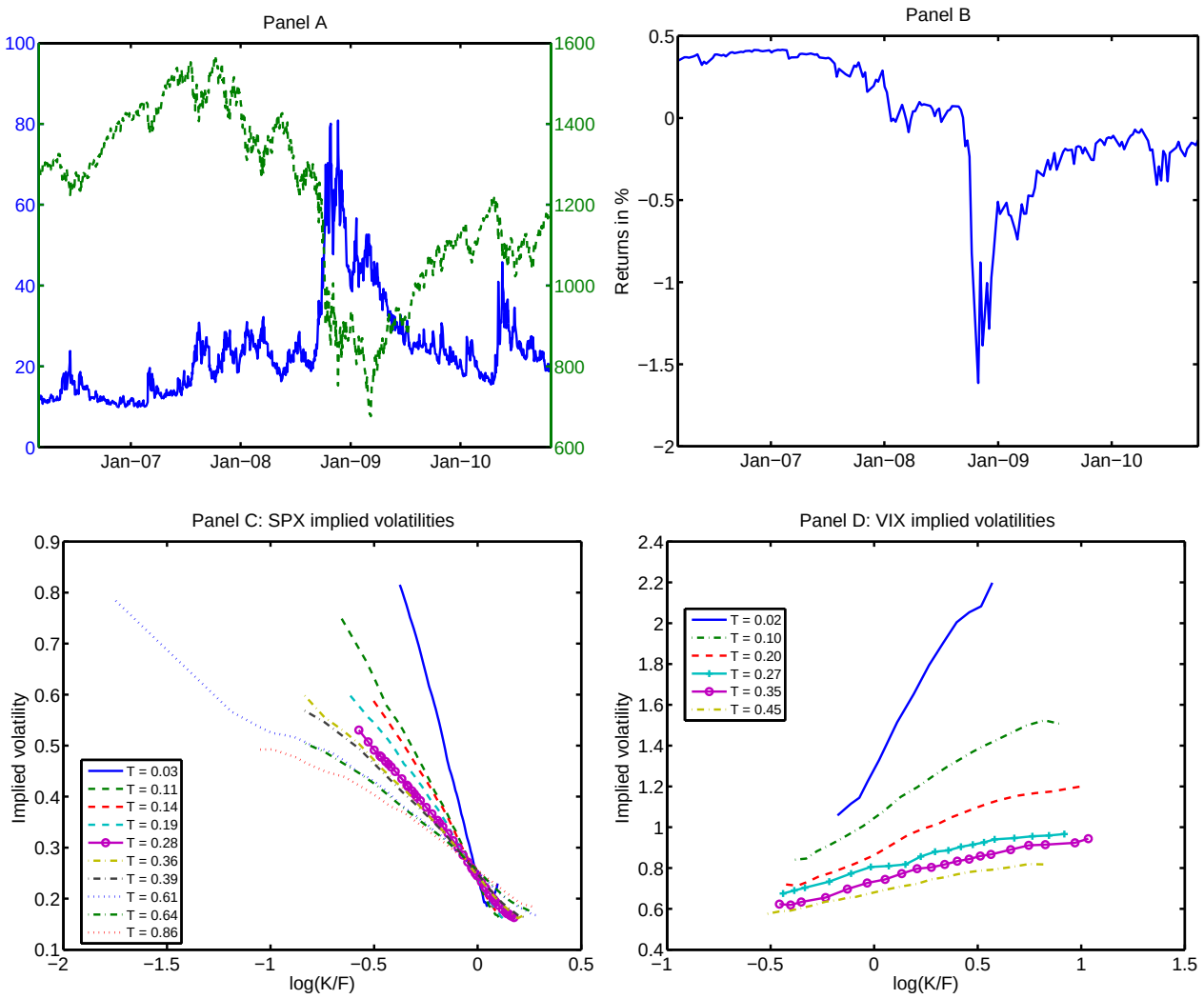

Figure 1: Panel A: Times series of S&P 500 (dashed curve) and VIX (solid curve) indexes from March 1st, 2006 to October 29th, 2010. Panel B: One-month S&P 500 and VIX expected returns implied by S&P 500 options with one month maturity. We use the method described in Bakshi et al. (2003). Returns are expressed in percentages. Panels C and D: Implied volatilities of S&P 500 options and VIX options on May 10 2010, as a function of log-moneyness $\log(K/F)$. The maturities T are quoted in years.

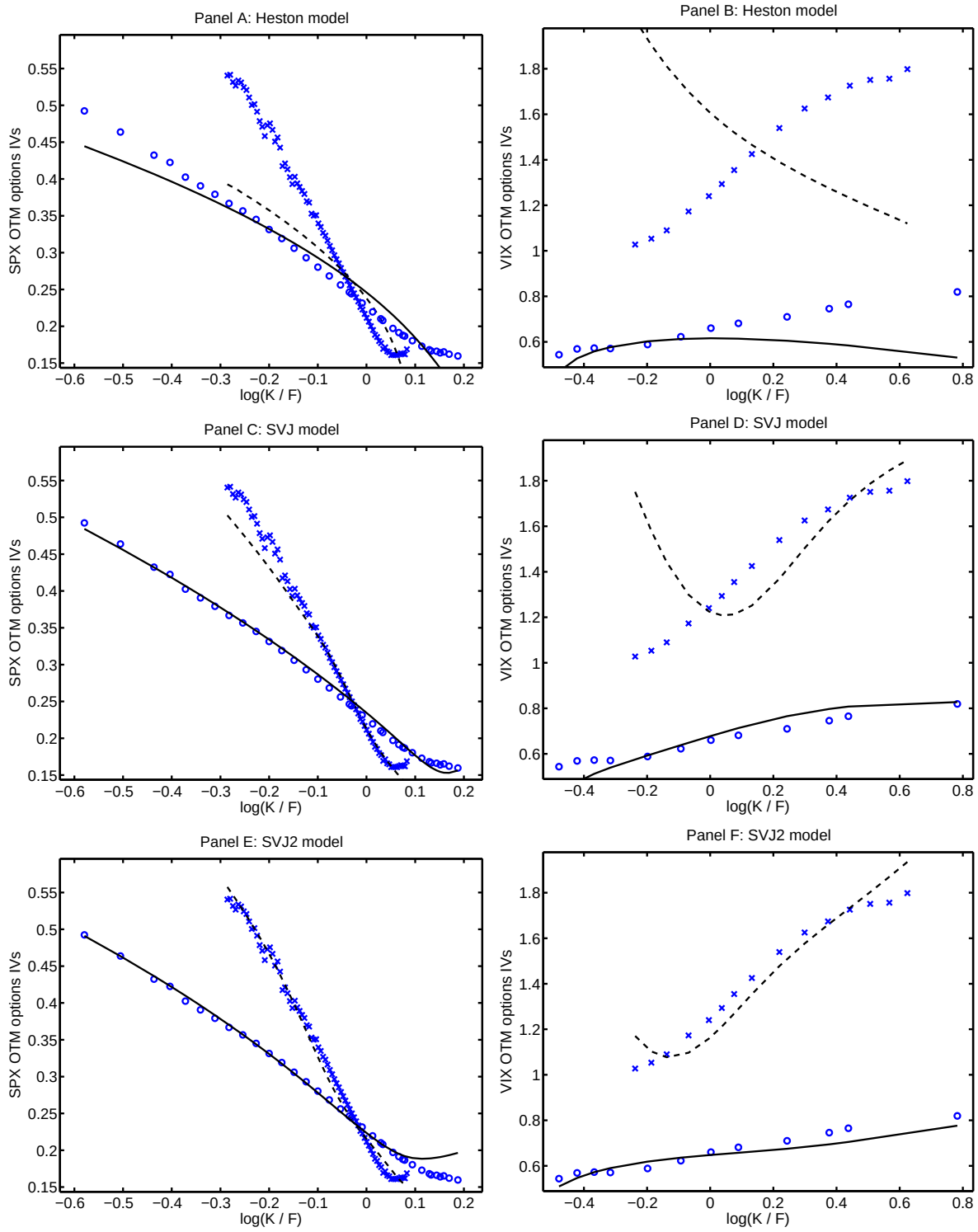

Figure 2: Market and model IVs on May 5, 2010, obtained by a joint calibration on the S&P 500 and VIX option market. Circles represent the market IV for $T = 0.05$ (S&P 500) and $T = 0.04$ (VIX). Crosses represent the market IV for $T = 0.3$ (S&P 500) and $T = 0.36$ (VIX). The dashed line corresponds to the model fit for $T = 0.05$ (S&P 500) and $T = 0.04$ (VIX) while the solid line corresponds to the model fit for $T = 0.3$ (S&P 500) and $T = 0.36$ (VIX). Panels A (S&P 500) and B (VIX) plot the model IVs based on the Heston model. Panels C and D display the corresponding results for the SVJ model, while Panels E and F do so for the SVJ2 model.

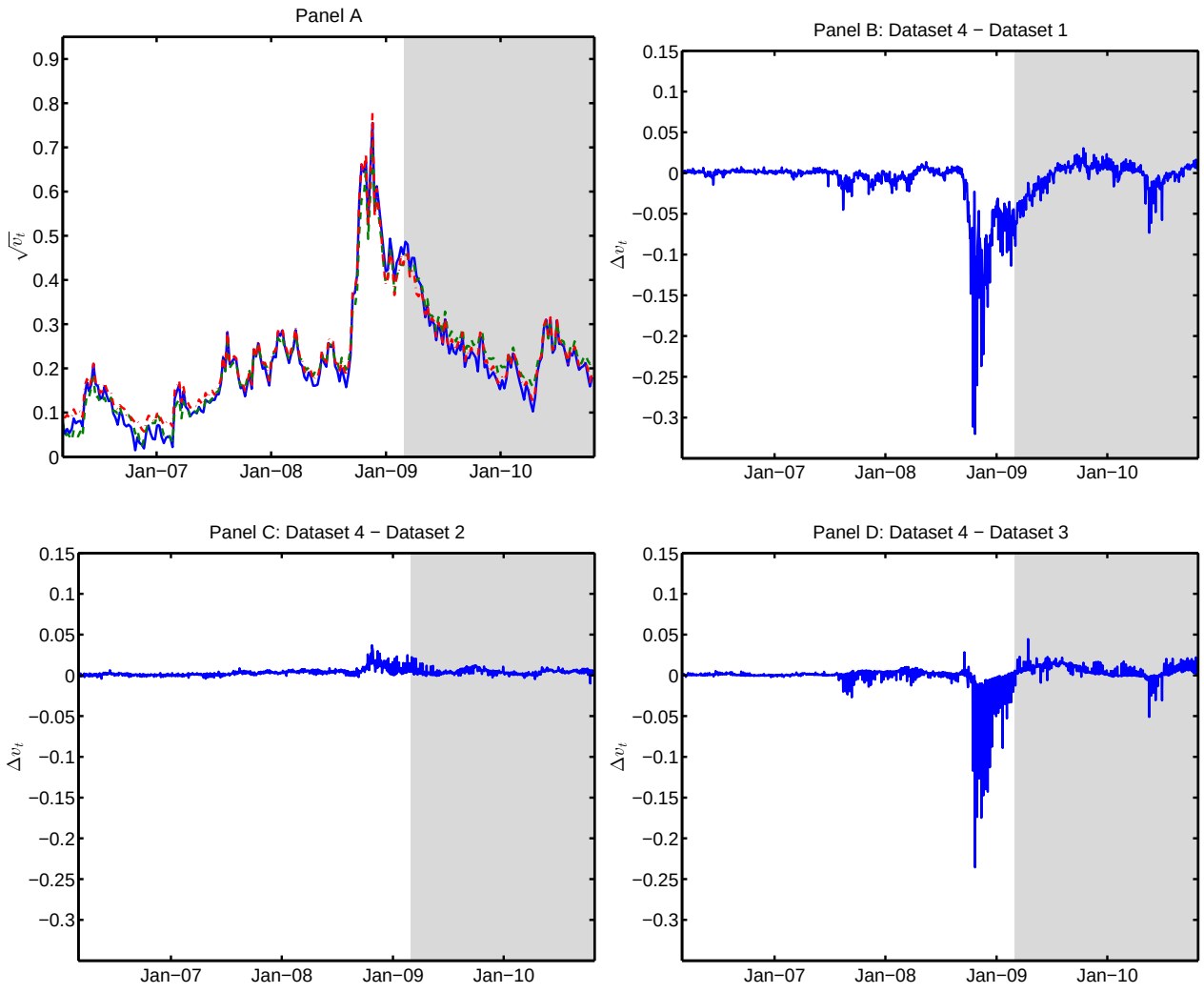

Figure 3: Panel A: Filtered trajectories of the latent volatility process v for the SVJ2 (solid line), the SVJ (dashed line), and the SV2 (dashed dotted line) using $\mathcal{D}_4$ from March 2006 to March 1st, 2009 (748 days). Panels B, C, and D: difference between the filtered variance of the SVJ2 model for dataset $\mathcal{D}_4$ and other datasets. The shaded part of the graph represents the out-of-sample period, from March 1, 2009 until the end of October 2010.

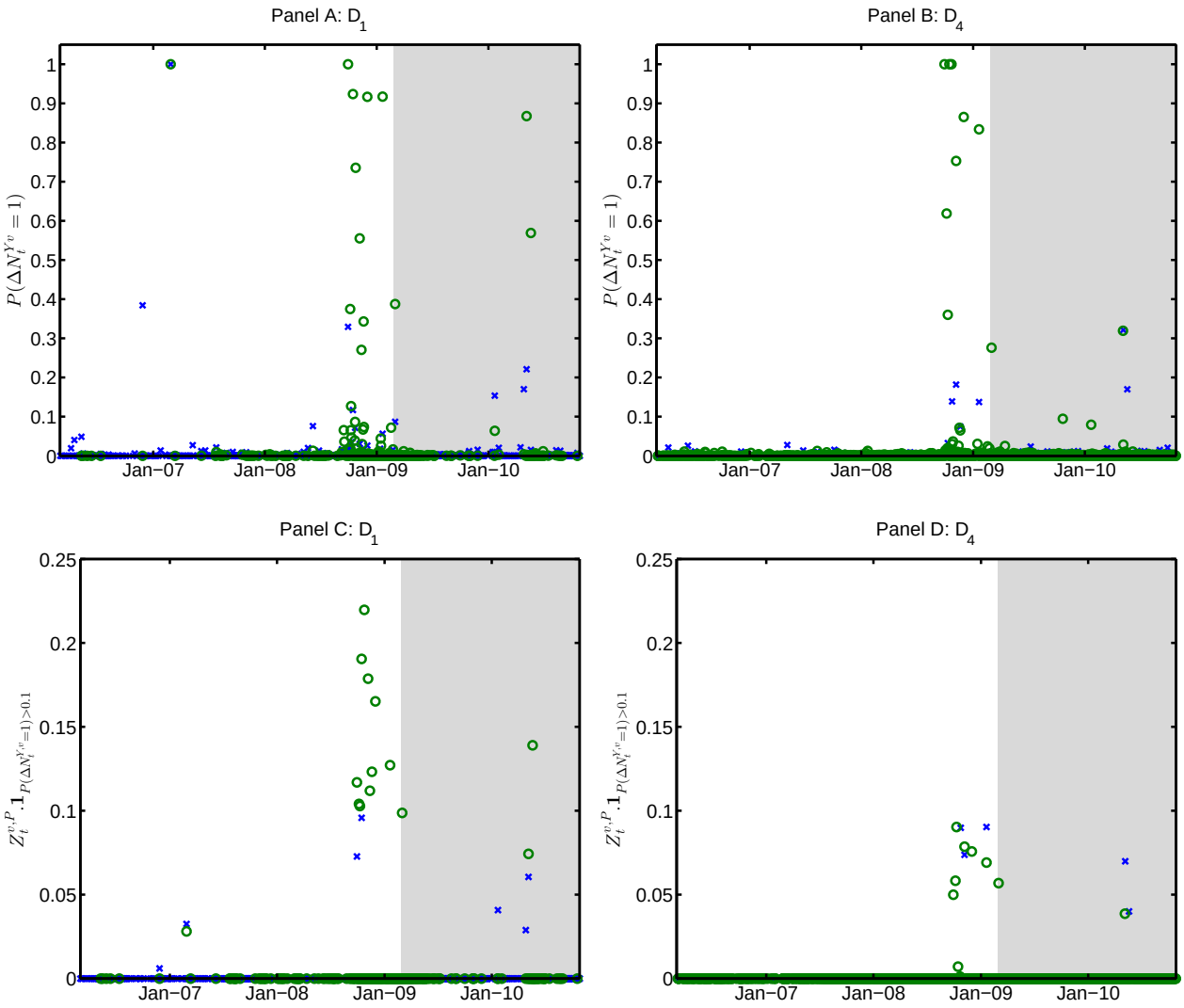

Figure 4: Filtered probability of jumps in the variance process v for the SVJ2 (crosses) and the SVJ (circles) for $\mathcal{D}_1$ (Panel A) and $\mathcal{D}_4$ (Panel B). The data covers March 2006 to March 1st, 2009 (748 days). In Panels C and D, we plot the filtered jump sizes in the variance process v when estimating the SVJ2 (crosses) and the SVJ (circles) models for $\mathcal{D}_1$ and $\mathcal{D}_4$, respectively. The shaded part of the graph represents the out-of-sample period, from March 1, 2009 until the end of October 2010.

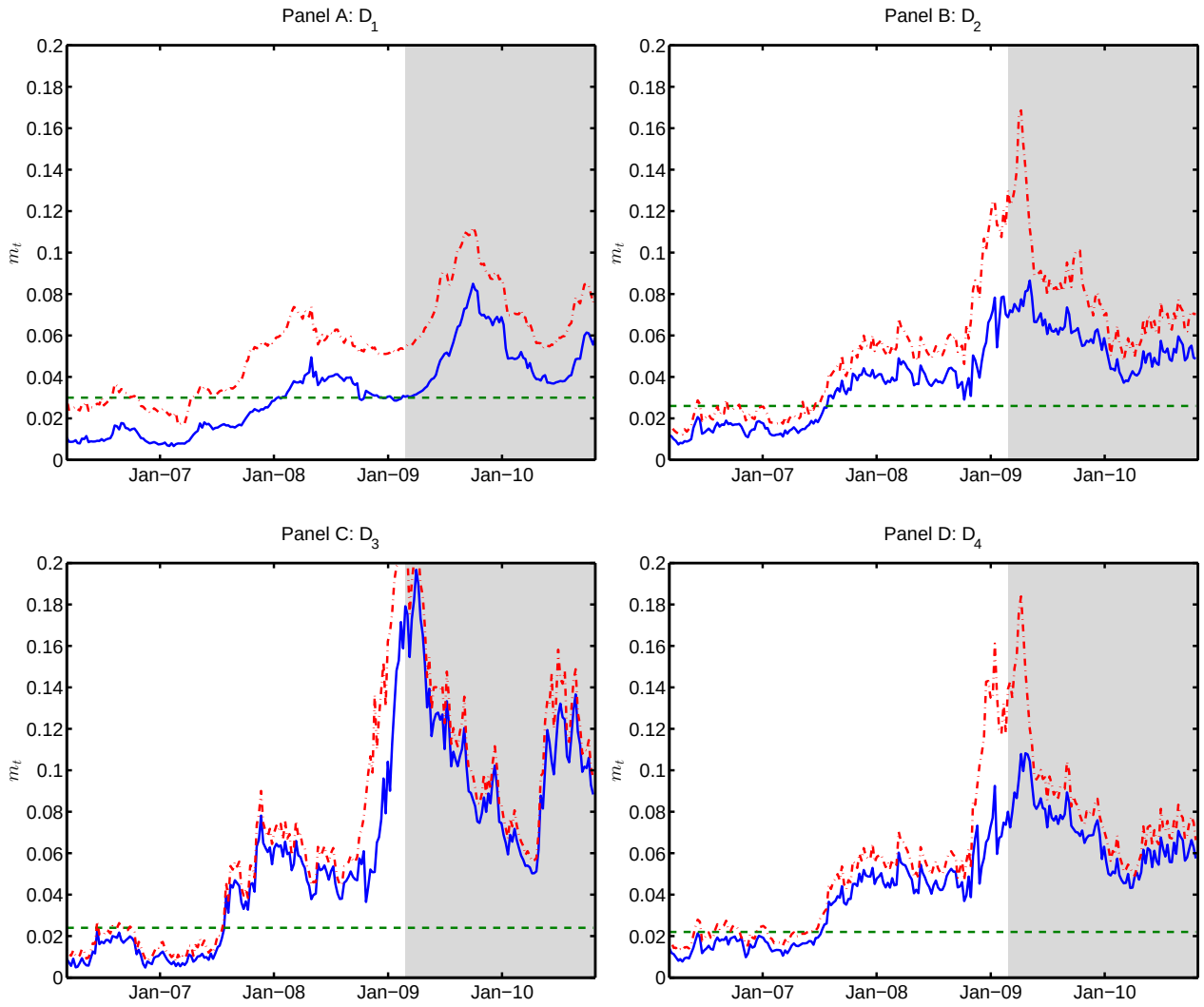

Figure 5: Filtered trajectories of the latent factor m when estimating the SVJ2 (solid line), SVJ (horizontal dashed line), and the SV2 (dashed dotted line) models over the different datasets $\mathcal{D}_1$ to $\mathcal{D}_4$ from March 2006 to February 2009 (748 days). The shaded part of the graph represents the out-of-sample period, from March 1, 2009 until the end of October 2010.

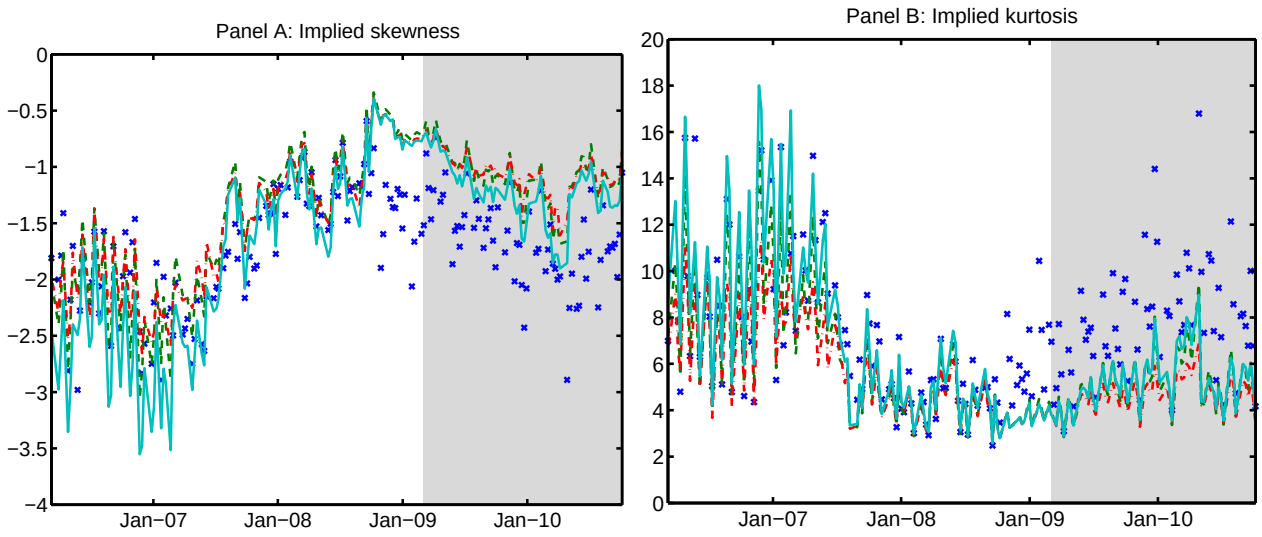

Figure 6: One-month risk-neutral skewness and kurtosis of the distribution of returns implied by 1-month S&P 500 options prices when estimating the SVJ2 (solid line), the SVJ (dashed line), and the SV2 (dashed dotted line) models using dataset $\mathcal{D}_4$ from March 2006 to February 2009 (748 days). We use the method described in Bakshi et al. (2003). The shaded part of the graph represents the out-of-sample period, from March 1, 2009 until the end of October 2010.

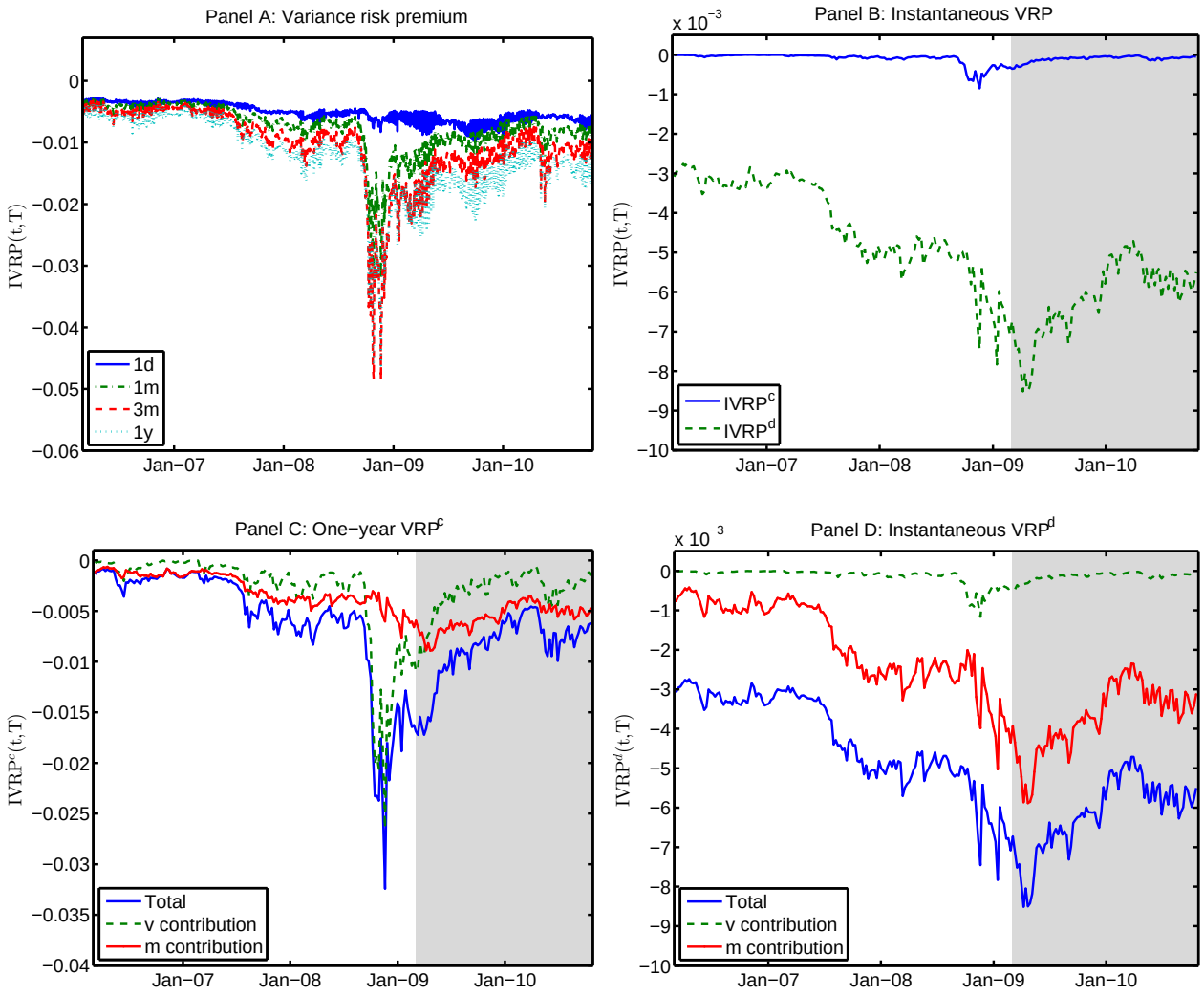

Figure 7: Integrated variance risk premia (IVRP) when estimating the SVJ2 model using the full dataset $\mathcal{D}_4$. Panel A plots the IVRP for different maturities. Panel B decomposes the instantaneous IVRP into its continuous and discontinuous components. Panel C plots the contribution of the latent factors v and m to the continuous component of the 1-year IVRP. Panel D shows the contribution of the latent factors v and m to the discontinuous component of the instantaneous IVRP. The shaded parts of the graphs represent the out-of-sample period from March 1, 2009 until the end of October 2010.